Combined systems approaches to understand host-pathogen interactions

Kozlowska, Justyna

Awarding institution:
King's College London

KING'S College LONDON

King's College London

PhD Thesis

Combined systems approaches to understand host-bacterial interactions

Author:
Justyna Kozlowska

Supervisor:
Dr. A. James Mason
Dr. Ken D. Bruce

A thesis submitted for the degree of
Doctor of Philosophy
in the Institute of Pharmaceutical Science
of the Faculty of Life Sciences & Medicine

Acknowledgements

I would like to take this opportunity to express my deepest appreciation to all the people who have helped me out over the last four years. I am fortunate enough to have the support of so many people and without them my project would not be where it is now. Some people helped on the project directly, while others gave me a push in the right direction. To every one of them I am eternally grateful for their help and support.

This work was supported by BBSRC CASE studentship awarded to Dr A. James Mason at King's College London and Dr Michael McArthur at Procarta Biosystems, so firstly I want to thank them, as without them this project wouldn't have happened. The most instrumental person for this thesis was my supervisor Dr James Mason, whose patience and guidance know no bounds and who constantly encourages me to go beyond my aims. I have worked with him for the last 5 years, first as a MRes project student, and now as a PhD student. I would also like to thank him for all the hours he has spent proof reading my reports, drafts of this thesis and also for our chats about life in general while having a drink every now and then. The work presented in this thesis owes much to his enthusiasm and careful guidance. I would like to thank also my second supervisor Ken D. Bruce for his advice and expertise and a great input in my publications and this thesis.

I wanted to thank the person I worked with on a daily basis for the first year of my PhD, Dr Louic Vermeer, for sharing so freely his time, knowledge and expertise which have been invaluable for the project even long after he left KCL.

In the IPS department I also worked with Dr Geraint Rogers who shared his time and expertise in working with RNA and later when we started a collaboration, which led to two publications and a deepening of my faith in my own abilities.

Thanks goes to great colleagues on the 5th floor of Franklin-Wilkins Building for stimulating conversations, dinners, evenings in the pub. I am particularity thankful for the time spent with Beatriz Padilla and Min Kim-our lunches, tea breaks, getaways and cycling trips, which will hopefully continue long after I leave King's College London. Also, a big thank you to Francesca Di Giuseppe, who once was a member of the lab and remains a great friend still. I also acknowledge Dr Andrzej Tkacz who I met while working with the industrial sponsor of my PhD and who has become a close friend always willing to cheer me up and give advice in times of doubt. A big thank you to Magda Swedrowska, who is a colleague, a flatmate and a friend and the last two years wouldn't be the same without her.

In the last 4 years I was fortunate to supervise many talented students whose work was included in my publications: Sara-Beth Amos, Kiran Shafiq, Nabila Rehnnuma and it was a real pleasure to work with them.

Also, a thank you to the NMR facility manager, Dr Andrew Atkinson, always so helpful and willing to answer NMR-related queries. In addition thank you to staff members in the King's Genomics Centre for help with processing and analysis of the GeneChips, especially Dr Matthew Arno, and staff members in the The Centre for Ultrastructural Imaging (CUI) who have also been extremely helpful and accommodating.

There are also other people who do not work with me but have been friends for many years and I would also like to thank them for their support - Hanna Rys, Bogdan Sciezka, Natalia and Alexei Matveyev and friends in Poland who stayed close despite the distance between us-Malgorzata Myczek, Monika Pietryka, Piotr Guzinski, Kasia Olczyk and Slawomir Michalowski. My flatmates and ex flatmates who put up with my moods, have supported me in my endeavours and provided a welcome distraction from work, particularly Gemma Glover, who remains a great source of inspiration despite living 10,000 miles away from London.

I would like to thank my mum, my grandmother and sister for supporting me for the last 28 years. It is through their encouragement that I have made it through all the steps to reach this point in life.

In summary, I would like to thank everyone for putting up with me for the past four years. I hope that this project will make a real contribution to the field and I hope that everyone that reads this thesis in its final shape finds it useful in their work.

Abstract

Systems biology studies are becoming increasingly important as the need to study organisms in a holistic manner, instead of looking at processes in isolation, is being recognised. This is especially true for the study of host-pathogen interactions where the responses from bacteria are complex and overlap extensively. This thesis explores the application of ^{1}H NMR metabolomics to the study of bacteria-host interactions and seeks to identify its strengths and weaknesses with a view to integrating this technique into a combined approach that can provide an unprecedented, sophisticated understanding of host-pathogen interactions that we believe is intractable by other methodologies.

The BBSRC CASE studentship that supported this work was awarded in conjunction with Procarta Biosystems Ltd who have produced a new generation of antibiotics with a novel mechanism of action. The final objective for the studentship therefore, was to develop a validated systems approach capable of defining the mechanism of action of this new class of antibiotics; transcription factor decoys (TFDs). By understanding how the target bacteria respond to antibiotic threats, the future development of new targets, delivery systems and formulations can be undertaken in a rational manner. The thesis builds towards this ultimate goal in three stages by showing, in a stepwise manner, how three increasingly complex scenarios can be interrogated by NMR metabolomics, either as a standalone technique or in combination with biophysical or genomic tools.

In the first stage we investigated how the growth of different *Pseudomonas aeruginosa* isolates from Cystic Fibrosis patients might influence airway secretions. Growth and NMR analysis of the spent media was technically challenging, highlighting the need for improved data pre-processing techniques and experimental design. Nevertheless multivariate analysis of changes in spent media composition could be related to univariate clinical measures of respiratory dis-

ease.

In the second stage we undertook a murine faecal microbiome study to show how different gut microbial communities affect the host gut metabolome. Faecal pellets were extracted into aqueous buffer and ^{1}H NMR spectra obtained in the solution state. Clear differences in the amino acid and short chain fatty acid complement of the mouse gut were related to divergence in the gastrointestinal microbiota in the mice. The study required comparison of two separate sets of multivariate data and showed how, with application of Hierarchical Cluster Analysis, relationships between microbiota can be simplified to generate hypotheses that can be tested using metabolomic approaches. In this study the metabolomic technique was capable of identifying a link between divergence of gut microbiota and the nutritional performance of the mouse gut.

In the third and final stage, we investigated whether a whole organism view could provide a bacterial perspective to enable a better understanding of how bacteria respond to antibiotic challenges. Here we combined ^{1}H NMR spectroscopy, now of solid, bacterial cell samples (using high resolution magic angle spinning), with electron microscopy and transcriptomics to characterise the effect on *Escherichia coli* of four structurally and physically related antimicrobial peptides with suspected differences in their mechanisms of action. Bacterial responses characterised by the NMR metabolomic study could be detected at sub-lethal antimicrobial peptide concentrations and were qualitatively different according to the antimicrobial peptide. The technique was sufficiently sensitive and high-throughput to allow both a range of antimicrobial peptide concentrations to be probed as well as the bacterial response to be followed over time. Using the NMR technique to identify optimal conditions for GeneChip experiments allowed the antimicrobial peptide mechanism of action to be inferred from analysis of the ontological profile of those genes whose expression is altered in response to the antibiotic challenge. This study provided a fresh, novel perspective for previous functional and biophysical studies and shows that, with better integration with transcriptomic and other systems data, ^{1}H NMR metabolomics will have considerable value in the study of host-pathogen interactions.

Contents

List of Figures

List of Tables

Glossary

B. cenocepacia *Burkholderia cenocepacia.*

E. coli *Escherichia coli.*

P. aeruginosa *Pseudomonas aeruginosa.*

S. aureus *Staphylococcus aureus.*

CFTR cystic fibrosis transmembrane conductance regulator.

COSY homonuclear correlation spectroscopy.

CPMG Carr-Purcell-Meiboom-Gill spin-echo NMR sequence.

ENAC the epithelial sodium channel.

HSQC heteronuclear single-quantum correlation spectroscopy.

icoshift interval correlation shifting.

JRes J-resolved NMR spectroscopy.

MRSA methicillin-resistant *Staphylococcus aureus*.

NIPALS nonlinear iterative partial least squares algorithm.

NOESY nuclear Overhauser effect spectroscopy.

PRESS predicted residual sum of squares.

PTW parametric time warping.

QSAR quantitative structure-activity relationship.

VPdtw variable penalty dynamic time warping.

WATERGATE water suppression through gradient tailored excitation.

Acronyms

(p)ppGpp collective term for ppGpp and pppGpp.

AMPs antimicrobial peptides.

CF cystic fibrosis.

CL cardiolipin.

COW correlation optimized warping.

CSA chemical shift anisotropy.

EPS exopolysaccharide.

FEV_1 forced expiratory volume in 1 s.

FT Fourier transform.

GPL glycerophospholipid.

HCA hierarchical cluster analysis.

HR-MAS NMR high resolution magic angle spinning NMR.

HSD honest significant difference.

HSP heat-shock protein.

HSR heat-shock response.

ICE integrating conjugative element.

IM inner membrane.

IR infrared spectroscopy.

KDO 2-keto-3-deoxyoctonic acid.

LOOCV leave-one-out cross-validation.

LPS bacterial lipopolysaccharride or endotoxin.

LTA lipoteichoic acid.

MHz megahertz.

MIC minimum inhibitory concentration.

MMR mismatch repair.

MS mass spectrometry.

NMR nuclear magnetic resonance.

OM outer membrane.

OPLS orthogonal PLS.

OPLSDA orthogonal PLSDA.

OxyR hydrogen peroxide-inducible genes activator.

PBP penicillin binding protein.

PC principal component.

PC phosphatidylcholine.

PCA principal component analysis.

PE phosphatidylethanolamine.

PG phosphatidylglycerol.

PLS partial least squares regression or projection to latent structures.

PLSDA PLS discriminant analysis.

Pol polymerase.

ppGpp guanosine 3‘-diphosphate 5‘-diphosphate.

pppGpp guanosine 3‘-diphosphate 5‘-triphosphate.

PQN probabilistic quotient normalization.

Psp phage-shock protein.

RF radio frequency.

RNA ribonucleic acid.

RNS reactive nitrogen species.

ROS reactive oxygen species.

SM sphingomyelin.

SoxR redox-sensitive transcriptional activator.

ssDNA single stranded DNA.

TCS two component system.

TMS Tetramethylsilane.

TMSP 2,2,3,3-D4-3-(Trimethylsilyl) propionic acid sodium salt.

Symbols

δ chemical shift.

ν frequency.

$\boldsymbol{I}$ nuclear spin quantum number.

σ sigma factor.

$\boldsymbol{B}_0$ static magnetic field.

Physical Constants

$\boldsymbol{h}$ Planck's constant 6.626×10^{-34} J.

1 Introduction

1.1 Introduction

The human body contains 10 times more microorganisms than its own cells and researchers now estimate that more than 10,000 microbial species occupy the human ecosystem [3]. Some bacteria are beneficial or even vital for a healthy population, whereas others are linked to disease and it is important to learn how this correlates with changes in the microbiome. The study of complete systems using approaches that provide a holistic view of host-bacterial, in particular host-pathogen, interactions is becoming an essential tool in the battle against antimicrobial resistance. Given the complexity of host-bacterial interactions a systems biology approach is regarded as necessary to move forward in our understanding of biological processes involved [4–6]. Host-bacterial interactions are considered dynamic, complex and multifactorial and a combination of advanced -omics technologies is required to access the level of detail necessary to resolve the key interactions in their entirety. Therefore, the broad aim of this work is to develop new ways of investigating how bacteria interact with their environment and host: how they cope with environmental changes that confer stressful stimuli such as a challenge by antimicrobial peptides, but also how bacteria influence their environment. In this chapter, and each of the results chapters, it is argued that host-bacterial interactions are complex and deriving meaningful biological information from such systems is challenging. This thesis suggests that metabolomics techniques (here based on NMR spectra of bacteria or their products) might be a cost-effective and high-throughput means of characterising these complex interactions. The key advantage of metabolomics over other -omics approaches is that it allows reduction of the overall complexity of the system, while retaining the ability to maintain an overview of the complete system. This is due to the fact that, in a bacterial cell, metabolites should be less abundant than genes and proteins [7]. The challenge in applying metabolomics techniques to host-bacterial interactions is to retain the ability to understand biological function while reducing the number of reporters from biological processes in the cell. Our ultimate goal is to combine two existing techniques, metabolomics and transcriptomics, in a novel approach that provides a platform to probe host-bacterial interactions yielding mechanistic understandings that are not tractable through other means. Our approach here has been to relate NMR metabolomic data to, first, univariate measures of patient outcomes, a simplified but multivariate analysis of gut microbiota and ultimately the highly multivariate data available from GeneChip. At each step the strengths and weaknesses of the NMR metabolomic method are

considered, as is the ability to effectively relate this data to a biological understanding of host-bacterial interactions.

Understanding the functional connections between genes, proteins and metabolites is one of the greatest challenges in the postgenomic era and -omics approaches are often applied simultaneously in an integrative approach coupled with improved performance of bioinformatics. Proteomics is a direct method of analysing the function of genes by systematic identification and quantification of its associated proteins. However, such studies are complex and the results not always discriminatory, since one gene can give rise to more than one protein through alternative splicing or post-translational modifications in eukarotes. Post-translational modifications can also lead to profound alterations in protein activities and characteristics [8]. Another molecular strategy, transciptomics, measures the expression level of mRNA in a given cell population. Although transcriptome and proteome analyses provide important aspects regarding physiology of a cell, much is still unknown and it has been argued that the generation of hypotheses through these -omics data alone, is incomplete and may lead to incorrect interpretations [8]. Recently metabolomics has emerged as an important tool in understanding biological systems. Metabolites (small organic molecules) are either the end products or currency of cellular processes, hence the metabolome level should be an indicator of the transcriptomic and proteomic changes and reflect the response of biological system to environmental changes [9]. Metabolomics aims to define the complete set of metabolites for a biological system at a specific time or a cellular state in a holistic approach and provides major advantages over other -omic platforms:

- metabolomic data is easier to manage, as metabolites are less abundant than genes, transcripts and proteins, therefore the complexity is reduced (there are 2,700 metabolites in *E.coli*, *ca.* 5,000 genes with ten times greater proteome size and a number of possible protein-protein interactions much exceeding this number [10])

- metabolites offer a better way of discriminating between cellular states / conditions as changes in concentrations of metabolites are faster than in proteins and transcripts and sometimes changes in proteins and transcripts do not result in pertinent changes to the phenotype or in metabolites [9, 11]

- with necessary infrastructure (having access to the equipment), analysis of the metabolome is more cost-effective and high-throughput than of the

proteome or the transcriptome

Regardless of the platform used, metabolomic data generation consists of five essential steps outlined in the next chapter (Chapter 2 on page 18):

1. Sample preparation
2. Raw data generation e.g. LC-MS chromatogram or NMR spectra
3. Data pre-processing e.g. spectral calibration, baseline correction
4. Multivariate data analysis and integration into biologically relevant information
5. Setting of the results into biological context

The importance of each of these steps are considered in the context of different applications, of NMR metabolomics to host-bacterial interactions, which each seek to obtain biologically relevant information from an overview of the main metabolic variables of each system.

1.2 The complexity of host-bacterial interactions, a role for metabolomics?

In the 'damage-response framework' [12], which outlines the contributions of both the pathogen and the host in the process of pathogenesis, the pathogen is defined as a microbe that has the ability to cause host damage, virulence is the relative ability of the pathogen to cause host damage, whereas the virulence factors combine all the properties of the pathogen that can damage a susceptible host. Often, it is properties of the host at a given time that will determine whether bacteria are pathogenic or not, thus different hosts can respond differently to the attack of the same pathogen. Moreover, the same host can have a differing susceptibility to the same pathogen at different times. Generally, there are four possible outcomes of a bacterium interacting with a host: commensalism, colonisation, persistence or disease depending on the ability of the bacterium to avoid, manipulate or disarm host defence responses. Furthermore, at any given moment of host-bacterial interaction each of those states can occur at the same time [12]. In order to survive stress conditions, including the host environment, and maintain homoeostasis, bacterial cells respond through coordinated changes in gene expression and cellular metabolism via highly integrated regulatory networks. Bacteria

have evolved to constantly monitor and counteract harmful environmental insults and are equipped with stress sensing regulatory systems, which operate predominantly at the level of transcription initiation [13]. This can be achieved by regulation of gene expression or by using alternative sigma factors (σ), which are specificity proteins that can redirect ribonucleic acid (RNA) polymerases and initiate transcription from alternative promoters by substituting the primary σ [14]. Also, the concentrations of certain metabolites can affect gene expression, which in turn can regulate metabolic activity [15] (Figure 1.1 on page 9). For example low levels of tryptophan in the environment induce the expression of genes responsible for production of this amino acid. According to Vinson *et al.* [16] despite *Escherichia coli* (*E. coli*) responding to the environmental changes by regulating the level of enzyme expression to keep metabolite levels stable, messenger RNA and protein levels do not change in response to most gene disruptants. Changes in the structure of the metabolic network keep the levels of most metabolites stable. This is supported by another study [17], which reveals emergence of an alternative pathway in central carbon metabolism in response to 12 different nutrient conditions. It did not require changes in gene expression, but rather relied on the accumulation of an intermediate metabolite. This study demonstrated the robustness of bacterial metabolic pathways. Brooks *et al.* [18] on the other hand postulates, using their data-driven models, that bacteria can tailor transcriptional responses to environmental changes, despite a modest number of regulators. It is still to be explored, however, to what extent the changes of metabolite concentrations and transcriptional response are related [15, 19] and weather metabolomics and transcriptomics approaches can be combined to gain insight into the coordination of these responses and their dynamics in the context of host-bacterial interactions. It is also unknown whether metabolomics can emerge as a stand-alone technique for this purpose.

1.2.1 The influence of host-bacterial interactions on patient outcomes, a role for metabolomics?

Bacteria are able to mount various stress responses to survive selective pressures from their environment. Pathogenic bacteria are also capable of deploying virulence factors that allow them to successfully colonise host organisms [12, 20]. The immune system of the host organism poses many barriers for bacteria and highly sophisticated virulence factors are required to facilitate entry, inactivate

host defence systems and provide recovery from the damage suffered by bacteria from host defences. Despite this large panel of stresses imposed by host cells or the host organism, pathogens need to access new environmental niches to win competition with other microorganisms and sustain growth [20]. Colonisation of a host is beneficial for pathogenic microorganisms as compared with free living bacteria, as abiotic conditions such as temperature, pH, pressure, osmolarity, humidity among others as well as supply of nutrients is constant [12, 20]. Mounting stress responses and activation of suitable virulence factors are often interconnected and, in contrast to specific environmental stress responses, much more complex. Bacteria are under assault from many different stressors simultaneously with even more signals to take into account in case of infection when the host innate and adaptive immune responses are active. This hinders the direct analysis of host-pathogen interactions using traditional, reductionist approaches.

Bacterial responses to host defence systems can be divided into defensive and offensive [13]. Defensive responses facilitate survival when faced with challenging conditions. Offensive mechanisms are represented by the virulence factors including an array of toxins, adhesins and invasins essential for active crossing of barriers in a host organism and elimination of its defence mechanisms. When considering defensive mechanisms, there are two ways in which bacteria can protect themselves from stress conditions and increase their chances of colonisation and proliferation: biofilm formation and modifications to bacterial surface structures. Pathogens can modify themselves and their unique antigens or disguise them by coating their surface with various host proteins, which results in inhibition of opsonising antibodies and any subsequent reactions such as phagocytosis or complement activation [21]. Another variation of this defensive mechanism is modifications to the composition of bacterial lipopolysaccharide or endotoxin (LPS) in Gram-negative bacteria and secretion of extracellular material in form of a slime layer or a capsule. Biofilms are multicellular agglomerations, structured communities where bacteria of distinct physiological states [22] attach to each other as well as host surfaces or abiotic surfaces and are embedded in a slimy extracellular matrix composed of proteins and/or carbohydrates [23]. Various persistent and destructive inflammatory processes are attributed to bacteria living as biofilms, since they can reach resistance levels 10-1000 times higher than planktonic bacteria, which are single-cells that may float or swim in a liquid medium [23]. The high persistence and decreased sensitivity to antibiotic treatment is caused by many factors. These include reduced metabolic activ-

ity, insufficient penetration of antimicrobial agents into biofilms and protection against phagocytosis, binding of effector molecules, activation of signalling pathways and other immune responses by the extracellular matrix. The pathogen is not recognised as potentially dangerous, therefore the host immune system cannot elicit the necessary response to eliminate it. This survival mechanism can be seen in cystic fibrosis patients, whose lung infections are dominated by *Pseudomonas aeruginosa* (*P. aeruginosa*). Secretions from *P. aeruginosa* potentially play a selective role in the overall microbiota composition in the airways and also affect CF lung function [24]. In our study (Chapter 3 on page 48) the impact of *Pseudomonal* secretions on key indicators of patient outcomes has been investigated. This study serves to highlight how a limited selection of univariate patient data can be related to the multivariate data obtained in a metabolomic investigation and potentially identify the key effects of bacterial colonisation.

1.2.2 Microbiome and metabolome, a causal relationship?

Deleterious effects can come also from products secreted by host cells to the extracellular environment and bacteria have to mount a survival response by up-regulating suitable stress responses and expression of virulence factors. These insults can include antimicrobial peptides [25] and other antibacterial agents such as lysozymes, bile salts, fatty acids, chloridhydric acid, but also resident microbiota. This can be seen in the specialised microbes that have evolved to survive the insults imposed on them by the range of niche habitats within the gastrointestinal ecosystem, such as Bacteriodetes and Firmicutes [26]. Such specialisation of microbiota to withstand the insults and adapt to the environment presents evolutionary advantages to the host. Lederberg [27] has emphasised the importance of having a broad, holistic view of such relationship. The existence of microbe and human is interdependent; the phenotype of colonising bacteria can influence the emergent properties of the community and this in turn can be a factor modulating e.g. nutritional extraction or susceptibility to certain diseases in some hosts [26, 28]. Metabolites, in turn, are effector molecules and can be the reason behind the associations between microbiome and conditions such as obesity, allergy etc., which are associated with alteration in the composition of the microbiota. Obese and lean humans differ in microbiota compositions [29]. It was demonstrated to modulate metabolism in germ-free mice when transferred with microbiota from obese and lean humans, highlighting the need for greater understanding of microbiota-host metabolic interactions. A recent publication

[30] demonstrates that microbiome disruption early in life, for example through exposure to antibiotics, can be linked to increased risk of obesity in adulthood. The gut microbiota has been also shown to be critical in the homeostasis of host metabolic and immune networks. When early microbial colonisation in mice was delayed, the gut associated lymphoid tissues (GALT) failed to develop, leading to persistent immune dysregulation [31]. Links between changes in microbiota due to antibiotic use, dietary changes, and other lifestyle differences and development of allergic diseases have also been shown [32]. Worldwide, sensitisation rates to one or more common allergens among school children are currently approaching 40 %-50 % [33]. In Europe, 87 million people suffer from allergies [33] and the issue and its link to gut microbiome received coverage in recent news [34].

Such a relationship between microbiome and metabolome was investigated in the mouse gut using NMR-metabolomics. We have examined faecal pellets and were able to detect differences in their composition which related to the divergence in the mice gut microbiota (Chapter 4 on page 83). The challenge here is to relate two sets of multivariate data (microbiome and NMR metabolome) to each other to understand how a change in gut microbiota might effect host responses manifested in diseases processes. Here the microbiome data is simplified by using a hierarchical cluster analysis that identified groups of mice whose gut microbiomes were closely related. This provided groups of mice whose gut faecal metabolomes could be compared to reveal how variation in gut microbiota can substantially alter the small molecule complement and nutritional balance of the animal.

1.2.3 The complexity of bacterial stress responses: Combining metabolomics and transcript profiling to investigate bacterial responses to antibiotics.

Biological systems are composed of functional units that interact at multiple levels via complex network of genes, proteins and metabolites. The interaction is regulated primarily by signalling molecules and enzymatic inhibition or activation reactions with transcription being controlled by levels of metabolic end products (Figure 1.1 on the following page). Therefore, in order to capture such complex and dynamic interplay, which spans so many regulatory levels, a systems biology approach is essential. To complete this task a combination of -omic

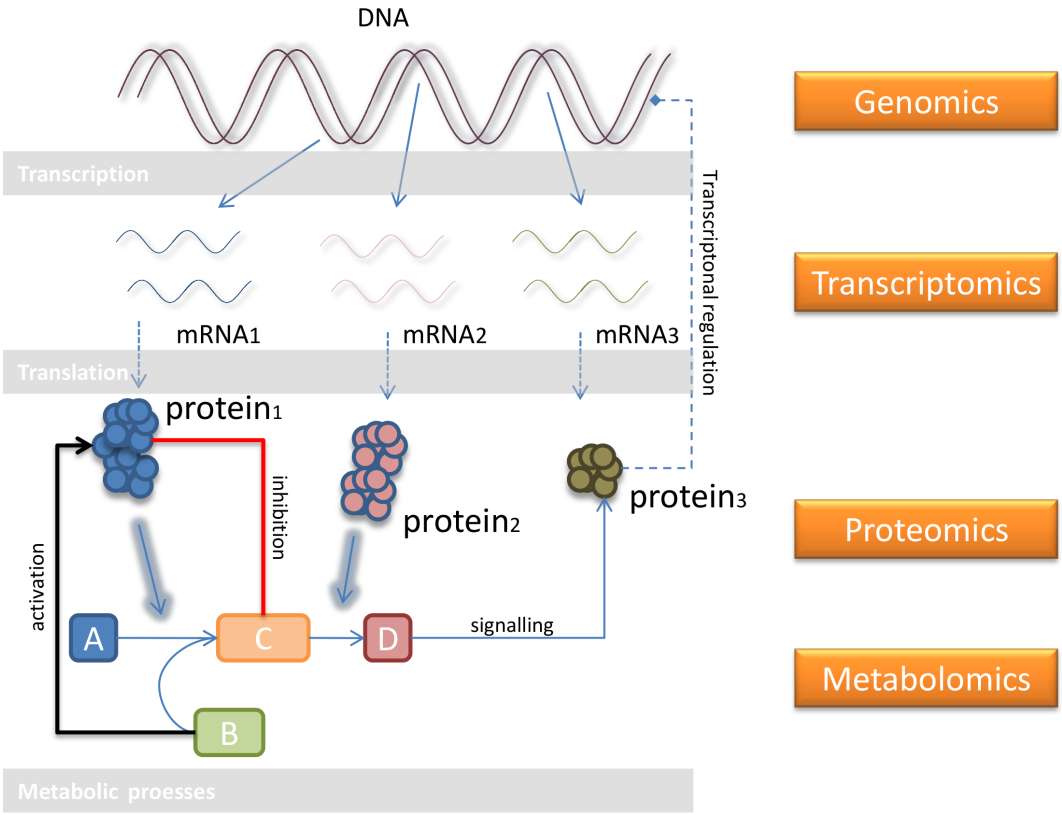


Figure 1.1: ***Systems biology level of cellular organisation.*** *Shows metabolic network and feedback regulation in response to developmental and environmental conditions. Modified from Carneiro, S.M A (2010). A Systems Biology approach for the characterization of metabolic bottlenecks in recombinant protein production processes. Ph.D. Thesis. University of Minho: Portugal.*

methodologies supported by biophysical tools were used, which investigate bacterial systems at different organisational levels. Combined -omics strategies can increase understanding of host-pathogen interactions by identifying biomarkers, giving insights into metabolic and genetic network regulation and identifying pathway bottlenecks and nutritional requirements. This knowledge will inform rational development of prevention and control measures. However, to be able to interpret such data and give it biological meaning, it is essential to differentiate between different bacterial responses and the possible outcomes of host-pathogen interactions at the transcriptome and metabolome level.

Bacterial stress responses can be classified into those induced by nutrient limitation (general stress response, the stringent response), cell damage or exposure to antibiotics (envelope stress, DNA damage, oxidative stress) or abiotic stress (heat stress, osmotic stress). Each of these types of bacterial stress response overlap and more than one stress response can be induced by the same stimulus

and the same alternative sigma factor can be induced independently in different types of stress response.

The general stress response is initiated by activation of alternative sigma factor RpoS (ς^{38}) in response to starvation and growth inhibition, but also growth rate reduction, extreme temperature changes, high osmotic pressure and low pH [35]. The RpoS regulon overlaps extensively with other global response networks and directly or indirectly controls over 200 genes in *E. coli* [36]. Another aspect of this response is induction of DNA polymerase (Pol) IV and Pol V. The activity of Pol V is tightly controlled and targeted in *E.coli* in order to prevent spontaneous mutations which are usually detrimental [37], whereas Pol IV activity leads to an increase in the spontaneous mutation rate of stressed cells. RpoS has also been proposed to be responsible for down-regulating enzymes essential for mismatch repair (MMR) [35]. Therefore, under stressful conditions bacterial cells are primed for genetic change and accumulate the mutations that increase their chances of survival via mutagenic phenomena referred to as adaptive mutation [36].

The stringent response is the second type of response triggered by nutritional limitation and amino acid starvation and its initiation is mediated by the RelA/SpotT enzymes [38]. It is characterised by accumulation of guanosine 3‘-diphosphate 5‘-triphosphate (pppGpp) and guanosine 3‘-diphosphate 5‘-diphosphate (ppGpp) nucleotides, collectively referred to as (p)ppGpp. The result of (p)ppGpp accumulation is the arrest of ribosomal translation and, in *E.coli*, (p)ppGpp mediates this process by inhibiting replication initiation [38] thereby shifting the expression of genes responsible for cell growth to expression of stress-related genes. Little is known about the impact of the (p)ppGpp-induced stress response on cellular metabolism, but it plays a key role in the adjustment of the bacterial growth rate, therefore continuously balancing between growth and survival. It facilitates the RpoH (below) and RpoS responses and multiple studies suggest that it may be involved in many other cellular responses such as DNA damage, osmotic, oxidative or SOS stresses [39].

Bacteria have evolved to have many ways to minimise lethal consequences of exposure to DNA damaging agents. The SOS response is triggered by DNA damage and is mediated by the LexA/RecA regulon [35]. Regions of single stranded DNA (ssDNA), a result of DNA damage or DNA repair, are recognised and com-

plexed with RelA, a bacterial recombinase, which in turn induces cleavage of the LexA repressor. Cleavage of the LexA repressor results in increased activity of enzymes essential for DNA repair, synthesis and recombination and tolerance mechanisms including DNA polymerases-Pol IV (above), Pol V, encoded by the *umuDC* genes, and Pol II, which is thought to be responsible for rescuing stalled replication forks [35]. In total, 42 genes have been identified to be regulated by LexA-RecA regulon including genes involved in efflux pumps, channels, membrane function [40]. Other deleterious conditions activating SOS response are exposure to antimicrobials, sublethal/lethal pH and hydrostatic pressure changes, oxidative stress, metabolic intermediates and nutrient limitation [40]. Recent studies revealed a network of genes in *E.coli* that play a role in modulating the response to DNA damage whose induction is independent from LexA/RecA regulon [41]. A study applying a metabolomic and transcriptomic approach to look at the effects of agents such as antimicrobial peptides targeting DNA could perhaps provide more insight into the regulation of those genes.

In the maintenance of homoeostasis in *E.coli* and other Gram-negative bacteria the first line of defence is the cell envelope and bacteria respond strongly to its perturbation. Those responses are highly regulated by alternative sigma factors and two component systems (TCSs). There are five major envelope stress responses in *E.coli*: sigma factor RpoE (ς^{E}), Cpx signalling system, Rcs (regulator of colanic acid synthesis), phage-shock protein (Psp), and Bae response [42]. (ς^{E}) is increased in response to a variety of conditions: heat shock, exposure to ethanol, treatment with antibiotics, mutations leading to misfolding of outer membrane (OM) and alterations in LPS. The Cpx regulon is triggered by alterations to the composition of the inner membrane (IM) and also alkaline pH, whereas the Rcs response is triggered by treatment with beta-lactam antibiotics and lysosyme, which inhibit peptidoglycan synthesis [43]. Envelope stress response pathways are therefore activated by generation and accumulation of misfolded periplasmic or membrane proteins and loss of OM integrity and the main aspects of the responses involve upregulation of genes encoding periplasmic chaperones and proteases that refold or degrade those proteins alleviating the stress response [43]. This helps in repair and biogenesis of the bacterial envelope and thereby restoration of cell surface integrity. Envelope stress response pathways also play roles in bacterial motility, colony and biofilm formation and virulence regulation [43].

Reactive oxygen species (ROS) and reactive nitrogen species (RNS) are generated in the course of metabolism, exposure to antibiotics and redox-active molecules and can also be a way of one bacterial species inhibiting growth of another [44]. They can cause damage to proteins, nucleic acids, and cell membranes and, to counter the effect of oxidative stress, bacterial cells constitutively express a number of antioxidant enzymes that detoxify the reactive oxygen species and repair the damage they caused. In addition to the SOS response, several dedicated responses to oxidative stress have been identified. Two major adaptive responses are regulated by the redox-sensitive transcriptional activator (SoxR) and hydrogen peroxide-inducible genes activator (OxyR) regulons in *E. coli* [45]. Moreover, exposure to ROS was shown to induce expression of genes required for induction of components of multidrug efflux system promoting resistance development [46]. Most of the cell death that occurs upon hydrogen peroxide exposure is thought to be due to DNA damage [47], whereas some of the DNA damage leads to miscoding and increased mutagenesis in aerobically growing cells [45].

The heat-shock response (HSR) is triggered upon activation of another alternative sigma factor-RpoH (ς^{32}) through a signal transduction system in response to a temperature shift above the normal growth range [35]. The consequence of the response is overproduction of proteins that increase the tolerance of bacteria to such thermal insult-heat-shock proteins (HSPs). The sudden increase in temperature leads to cytoplasmic stress and misfolded and unfolded proteins. Among HSPs are chaperones and proteases that aid elimination of such damaged proteins. Furthermore, in UV mutagenesis, molecular chaperone GroE, which is controlled by RpoH-regulon, protect Pol V from degradation therefore controlling the spontaneous mutation rates [48]. It has been suggested that GroE is important for Pol IV-induced adaptive mutation by protecting Pol IV from degradation [49]. The RpoH regulon can be induced by stimuli other than temperature shift: depletion of amino acids and carbon source, phage infection, challenge with antibiotics or heavy metals, DNA damage and oxidative stress [35].

Another form of stress, experienced by bacteria, that can alter cell structure, chemistry and physics is osmotic stress. Bacteria attenuate external upshifts and downshifts in osmotic pressure by accumulation (by *de novo* synthesis or uptake from the environment) of electrolytes or small organic solutes such as potassium ions, glutamate, proline or trehalose and release via mechanosensitive channels. The main aspect of the response is maintenance of turgor pressure within the

cells, by ensuring that the osmolarity of the cytoplasm is greater than the osmolarity of the medium [50].

To be able to thrive in a variety of habitats, bacteria had to evolve to respond to fluctuations in their environment. Bacteria have to be able to not only handle these nutritional and abiotic stresses, but also compete or share their environment with other organisms in their respective niches and hence may influence their own environment. Recent studies have focused on characterising those interactions and classifying them as cooperative, competitive or adaptive.

Therefore, in order to gain a comprehensive view of host-bacterial interactions and regulation mechanisms a method is needed capable of probing such complex responses.

1.3 Metabolomics

The metabolome is defined as the collective set of low molecular weight intermediates/metabolites in a biological system under particular physiological conditions [9, 11, 51]. Metabolomics studies the changes in the biological state of the cell, tissue, organ or the whole organism by non-biased identification and quantification of all detectable metabolites in the system [9, 11, 51] and the term was coined by Steve Oliver of University of Cambridge [9]. Its sister term, metabonomics, is defined as the quantitative measurement of the time-related multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification [52], however, the terms are often used interchangeably and the term "metabolomics" will be employed throughout this thesis.

Metabolites are both products and reactants of *in vivo* reactions and simultaneous characterisation of a number of metabolites provides a 'snapshot' of the dynamic state of an organism. Concentrations of metabolites play direct regulatory roles via feedback inhibition and allosteric mechanisms in rapid responses to metabolic flux changes. Therefore, metabolites are key biomolecules that control the cellular machinery, as elucidated by the central role they constitute in the regulome illustrated in Figure 1.1 on page 9. Moreover, the flux of metabolites is much faster (seconds) in comparison to turnover in proteome (minutes to hours) [53] and being downstream of both transcription and translation, metabolites might be better indicator of enzyme activity [54] making metabolomics more

suited for probing environmental perturbations and enabling the most current view of the state of an organism. Apart from scientific advantages, metabolomics has also major practical advantages-it is cost effective per sample and more high-throughput than other -omics, which makes it more suited for screening of a large number of conditions and their effects on the organisms, tissues or cells [55].

Despite being a relatively recent scientific development, metabolomics, is already emerging as an important tool in pharmaceutical discovery and development, the food industry, medicine, plant sciences and toxicology. It is widely used in target identification and investigation of the mechanisms of disease and toxicity and also characterisation of phenotypes and different metabolic states [56]. A PubMed search for the term 'Metabolomics' shows 1,559 entries in 2013 and 1,448 in 2014, which is over two-fold increase since 2009 largely due to recent experimental and technological advances. Metabolomic analysis, however, still faces a number of challenges due to the diverse nature and the amount of analytes present in each sample. This requires optimised protocols, particularly sample preparation methodologies, that take into account the chemical diversity and wide dynamic concentration range of the metabolites in a sample, and selection of a suitable analytical platform.

The metabolome comprises organic molecules such as amino acids, fatty acids, vitamins, carbohydrates and lipids, however interactions with inorganic species are also a subject of research [57]. The size of each metabolome varies greatly depending on the organism under study from over 2700 metabolites for *E. coli* [58] (http://www.ecmdb.ca/), 200,000 primary and secondary metabolites estimated for plant kingdom [51]. The Human Metabolome Database (HMDB, http://www.hmdb.ca) currently contains 41,818 metabolite entries and it grows each year.

Generally, the aim of metabolomic analyses is the generation of a list of metabolites with altered concentrations from which biological meaning can be derived. There are two general approaches that can be applied: the non-targeted and targeted approach. The aim of a non-targeted approach is non-biased identification of as many metabolites in a biological system as possible with no *a priori* knowledge of the sample nature and composition. Further multivariate analysis provides a list of molecular features discriminating between samples, allowing their classification according to treatment, disease, alteration and revealing po-

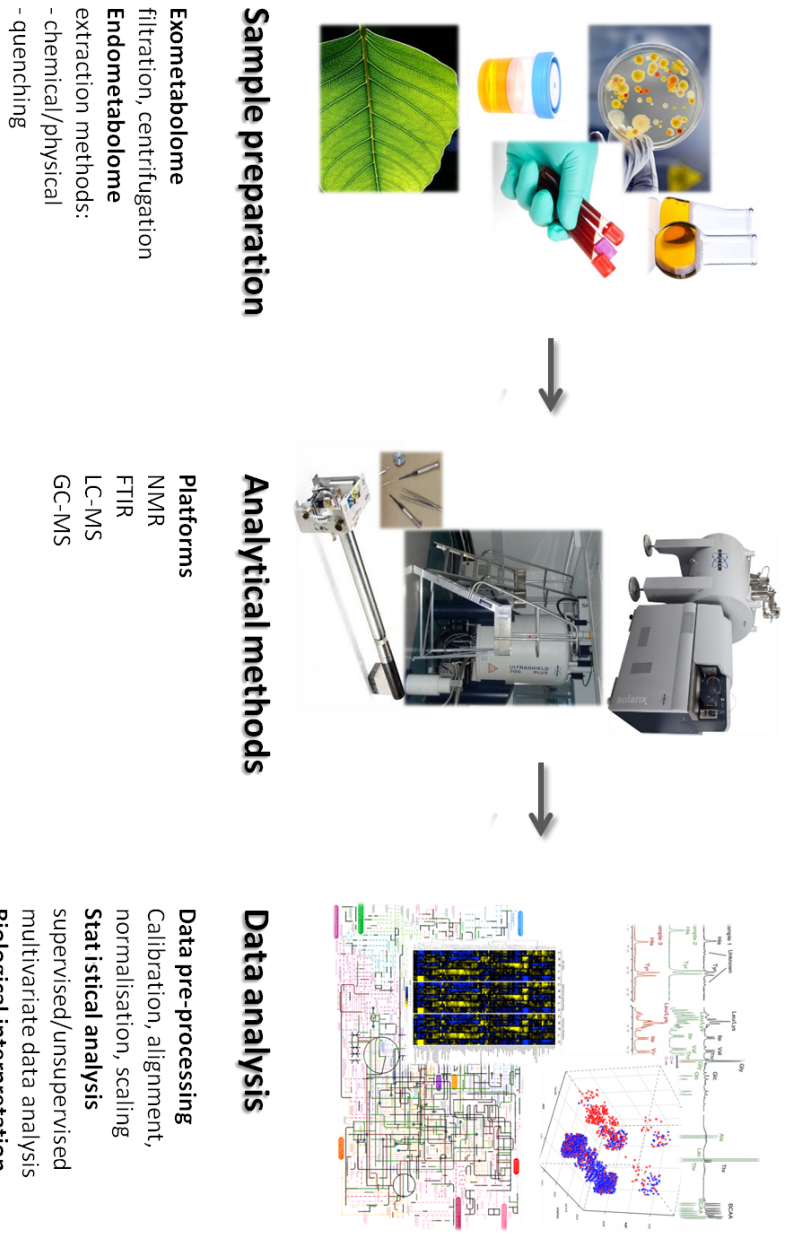


Figure 1.2: ***Metabolomics analysis workflow.*** *A typical metabolomics experiment involves sample preparation, data acquisition using MS and/or NMR and metabolite identification and quantification. Once a set of metabolites of interest have been identified, two types of tools can be used to gain biological insight into experimental results: (i) mapping and visualisation of pathways and (ii) statistical enrichment analysis of metabolite annotations.*

tential biomarkers for the condition. Targeted approaches, further divided into metabolic target analysis, metabolic profiling, metabolic fingerprinting [51] and metabolic footprinting [59], are used for qualitative and quantitative analysis of one of a few preselected compounds or classes of compounds in a sample, generally related to a specific metabolic pathway/metabolic reaction by selective analysis and specialised calibration methods. This approach is often used for drug screening and profiling of drug metabolic products [60].

Today, metabolomics applications range from finding a relationship between phenotype and related gene function, providing biomarkers useful for identification of early stage of certain conditions, understanding metabolic reaction networks and their *in vivo* regulation and predicting novel metabolic pathways to engineering of metabolite fluxes and measuring flux levels of metabolites of interest under varying conditions [51, 61–63].

1.4 Research objectives and thesis organisation

The first chapter of the thesis provides an introduction into the system-level overview of biological systems and why it is becoming a necessity to understand the multilevelled repertoire of host-pathogen interactions in biological systems. The overview of bacterial stress responses provided is linked with the globally increasing problem of antimicrobial resistance to antibiotics and provides an introduction to antimicrobial peptides (AMPs) and metabolomics. This is followed, in the next chapter, by an introduction to the main concepts and bottlenecks in NMR metabolomics and an outline of the process from data acquisition to data analysis. In Chapter 3 liquid-state NMR has been applied to study the pathogen, *Pseudomonas aeruginosa*, and its different isolates colonising the lining of lungs of cystic fibrosis patients. It has been investigated how this pathogen affects the host, i.e. patients lung function. In Chapter 4 both host and bacterium have been investigated in the study of the composition of a gut microbiota in mice and how differences in gastrointestinal bacterial makeup is linked to divergence in the host metabolome. Chapter 5 describes the main project, where multiple -omic approaches have been combined to measure the metabolic response of *E. coli* to AMP challenge along with transcriptional response and determine the extent of correlation between them. The chapter also provides information on physicochemical characteristics of the peptides used in this study, which were previously investigated in detail with circular dichroism (CD) and fluorescence

assays [64]. The ability of bacteria to initiate and coordinate changes in gene expression and at the metabolome level as a response to varying environmental factors is essential for successful pathogenesis and maintenance of homeostasis. In order to gain fundamental insight into the molecular mechanisms governing events associated with those processes the following objectives were considered in this thesis:

- How to investigate host-bacterial interactions in a cost efficient, reproducible and high-throughput manner? What are the strengths and limitations of such an approach?
- How to study antimicrobial mechanisms of action of novel therapeutics in a cost efficient, reproducible and high-throughput manner?
- Is the method sufficiently sensitive to distinguish between different bacterial responses to sublethal concentrations of structurally and physically related AMPs and how does this manifest in the metabolome?

2

NMR and data analysis in metabolomics: from data acquisition to pattern recognition and classification

2.1 Metabolomics workflow

2.1.1 Sample preparation

The key to a successful metabolomic experiment is the generation of high-quality biological samples. The type of sample is dictated by the character of the experiment, the target metabolites under study and the sensitivity of the method used. In metabolomics one can study intracellular (endometabolome) and extracellular (exometabolome) metabolites. Analysing extracellular metabolites provides information on what cells excrete into the intracellular matrix, particularly at suboptimal growth conditions, and/or what they fail to assimilate from their surroundings. While extracellular metabolites present in the extracellular medium can be filtered or centrifuged to separate them from the cells, the endometabolome sample preparation can involve considerable sample manipulation for some metabolomic platforms (Figure 1.2 on page 15). This is very important as, in order to provide meaningful results, it is crucial to minimise the experimental variation that may arise from different stages of sample generation and processing before the biochemical analysis. Heterogeneity and chemical complexity of metabolites is the main challenge in the field, since sample preparation methods must be non biased towards any group of metabolites be they volatile, non volatile, polar, semi polar or non polar and at widely differing concentrations. Generally this step involves quenching of metabolism and application of time-consuming and sometimes inaccurate extraction and separation procedures. Other factors that can greatly influence the reproducibility of the results are time and method of sampling as well as the storage of samples. Nuclear magnetic resonance (NMR) spectroscopy sample preparation can also be problematic due to presence of various buffers and viscous compounds in the sample and the dynamic concentration range of the metabolites under the analysis. The relatively low sensitivity of the method requires a substantial amount of sample, which has to be taken into account when planning the experiment to ensure a sufficient number of sample replicates. Determining the sample size for metabolomic experiments is a very important aspect, however currently there are no standard ways of sample size estimation in metabolomics, due to the complex nature of such experiment [65, 66]. The rule of a thumb is that one needs 3 times more samples than the number of components, with components defined as "independent sources of variation in the data" [65]. Technical replicates inform whether an outlier sample is actually biologically different, rather than the result of system variability.

2.1.2 Platforms for analytical quantification

The main efforts in metabolomics focus on the development of analytical platforms for metabolome quantification. Mass spectrometry (MS) is the most commonly used technique [51] with its variations: liquid chromatography mass spectrometry (LC/MS), liquid chromatography tandem mass spectrometry (LC/MS/MS), gas chromatography mass spectrometry (GC/MS) and capillary electrophoresis coupled to mass spectrometry (CE/MS). Other metabolome analyses technologies include: infrared spectroscopy (IR) and thin layer chromatography (TLC). Next to mass spectrometry NMR spectroscopy is the main metabolomics analytical platform. Since the first NMR metabolomic study over 40 years ago [67], the approach has experienced significant advancements in both instrumentation and analysis methodology and has become recognised and used extensively in clinical and pharmaceutical applications for the analysis of biofluids and tissues and in diagnostics for identifying biomarkers and defining pathological status. The method is powerful enough to produce reproducible and detailed spectra that provide a linear analytical response at the molecular level. Coupled with a high dynamic range, this allows comprehensive identification of multiple components within the sample simultaneously as well as providing quantitative characterisation in a non-targeted and non-destructive manner [56]. The method is used typically for biofluids where it involves little or no sample manipulation and the acquisition of the data is rapid and noninvasive [62, 68].

2.2 NMR for metabolomics

NMR is a very powerful and versatile technique that emerged in the mid-1940s [69, 70] and since then has seen unparalleled growth as an analytical tool for determination of molecular composition, structure, probing of dynamics and molecular reactions in areas such as chemistry, biology, materials science, medicine and geology. It can be applied to samples in the solid, liquid as well as liquid-crystal and gas states [71]. The application of NMR to solid samples will be discussed in detail in Section 2.2.2 on page 24.

2.2.1 The physical background to NMR

Nuclei in magnetic fields

NMR spectroscopy is based on the application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms that posses spin. The nuclei of all atoms can be characterised by a nuclear spin quantum number (I). It can have a value of zero or greater than zero, with all values being multiples of $\frac{1}{2}$. Any nucleus with an odd atomic number or odd mass number such as ^{1}H or ^{13}C will possess nuclear spin and be therefore amendable to NMR observation. For atoms with I=0, NMR cannot be used, as they have no nuclear spin. When an external, static magnetic field (B_0) is applied, the nuclei align themselves in a specific number of orientations with respect to the static field. For a spin of magnetic quantum number I, there are $2I$+1 possible spin states, whereas for a spin-$\frac{1}{2}$, such as proton, there are two possible states: $+\frac{1}{2}$ and -$\frac{1}{2}$ and it corresponds to the popular image of a nucleus having two spins which align parallel (α or upper state) or antiparallel (β or lower state), denoting lower and higher energy states, respectively [72].

Signal in NMR

If energy, $h\nu$, is absorbed, the emitted RF signal from excited nuclear spins is detected as a time-dependent oscillating voltage that steadily decays as a result of spin relaxation and which, after applying the procedure of a Fourier transform (FT), gives a signal in the frequency domain, i.e. a line in the NMR spectrum. The differences between the two population states are in the order of 1 in 10^4, even when very strong field is applied, rendering NMR a relatively insensitive technique [72]. The nuclear spins precess around the magnetic field vector and the ν (Lamor frequency) of this precession (Lamor precession) can be measured and is given by the equation 2.1

$$\nu = \frac{\gamma B_0}{2\pi} Hz \tag{2.1}$$

where γ is the magnetogyric ratio, which is a constant for any given nuclide. The nucleus may absorb photons whose frequency matches the difference between the low-energy and high-energy spin states, δE, according to equation 2.2:

$$h\nu = E_{upper} - E_{lower} \tag{2.2}$$

where h is the Planck's constant.

Features of NMR spectrum

Because the atomic nuclei are present in a molecule at different positions, they experience different interactions with the surrounding atoms. Of particular importance is the presence of the electron clouds, because they shield the nuclei from the external magnetic field. A reduced field experienced by the nucleus reduces the precession frequency of the nuclear spin, while electron withdrawing groups deshield a nucleus. This effect is referred to as a chemical shift (δ) and is defined as:

$$\delta = \frac{\nu - \nu_{TMS}}{\nu_0} \tag{2.3}$$

where (ν - ν_{TMS}) is the frequency difference between the resonance of the signal of interest and the reference resonance, which for ^{1}H NMR is 2,2,3,3-D4-3-(Trimethylsilyl) propionic acid sodium salt (TMSP) in aqueous solutions or Tetramethylsilane (TMS) in organic solutions in units of hertz (Hz) and ν_0 is the operating frequency of the spectrometer expressed in megahertz (MHz). This is a dimensionless property, but the ratio $\frac{Hz}{MHz}$ ($\frac{1}{1x10^6}$), therefore the units of parts per million (ppm) are used. TMSP/TMS signal is set by convention to 0 ppm. The ppm value is device independent, as it is given relative to the spectrometer frequency. The ppm values based on different references provide a relative chemical shift scale, which is used to distinguish between protons in a molecule and other nuclei such as ^{13}C, ^{15}N or ^{31}P in the different amino acids providing valuable structural information. The chemical shifts can be then assigned to specific metabolites.

Instrumentation and detection sensitivity

The majority of NMR applications of clinical relevance, including metabolomics, use ^{1}H NMR. Since the vast majority of metabolites contain hydrogen atoms, the technique is relatively non biased towards particular class of chemicals unlike other methods discussed above. The sensitivity of NMR, however, is relatively low and can still present an obstacle when compared to other spectrometric methods. ^{1}H signal-to-noise depends on many factors according to equation 2.4, however, an improvement can usually be achieved by either increasing the signal intensity or reducing the background noise.

$$\frac{S}{N} \propto N A T_s^{-1} B_0^{\frac{3}{2}} \gamma^{\frac{5}{2}} T_2^{*} (NS)^{\frac{1}{2}} \tag{2.4}$$

where N is the number of molecules in the observed sample, A is a term that represents the abundance of the nuclide, T_s is the temperature of the sample and surrounding rf coil, T_2^* is the effective transverse relaxation time and NS is the total number of accumulated scans. The signal intensity depends on many of the properties of the nuclide involved, such as the natural abundance, the magnetogyric ratio and relaxation properties, which are independent of the instrumental design. ^{1}H has high magnetogyric ratio, nearly 100 % natural abundance and favourable relaxation properties making it suited for high resolution NMR spectroscopy [73]. The number of molecules in the observed sample volume, N, is directly related to the amount of sample available in the active volume region of the coil.

In the last decade, significant advances have been made to boost the performance of NMR spectrometers and increase the detection limit in NMR spectroscopy, which is generally in the low micromolar range for less crowded parts of the spectrum. The magnetic field strength is one of the factors and a significant advance has been made in the development of more powerful magnets that also improve spectral resolution, as well as processing capabilities of new instruments. Moreover, the introduction of cryogenically cooled probes for high-resolution NMR spectroscopy offers 3-4 fold improvement in detection sensitivity [74]. Cryoprobes not only reduce the thermal noise by lowering the temperature of the rf detection coil, but also improve the signal-to-noise by amplifying the signal detected via inbuilt preamplifier, also kept cooled [74]. Spectra of large molecules not only suffer from excessive signal overlap, but also from poor sensitivity, because of their slow tumbling rate, which results in fast relaxation time i.e. signal loss and effectively broader lines. The introduction of pulse field gradients allows to selectively detect only the signal of interest and/or discard signals of no interest improving the optimal dynamic range of the experiment. A significant improvement in sensitivity has been also possible due to application of solvent suppression techniques with excitation sculpting [75], water suppression through gradient tailored excitation (WATERGATE) [76], presaturation being commonly used. In NMR metabolomics the most common NMR pulses used for metabolomic profiling are 1D nuclear nuclear Overhauser effect spectroscopy (NOESY) with presaturation (1D NOESY-presat) [77] and Carr-Purcell-Meiboom-Gill spin-echo NMR sequence (CPMG) spin-echo NMR sequence [78], because of high quality water suppression they offer. Moreover, CPMG allows removal of the broad resonances associated with macromolecules and slow tumbling compounds, by refocusing the

signal, therefore enhancing signal from low molecular weight metabolites. Also, the 2D ^{1}H-^{1}H J-resolved (JRES) NMR pulse sequence with water presaturation is sometimes used as it allows generation of much less congested spectrum via the JRES projection (p-JRES). JRES projections are effectively a proton-decoupled 1D ^{1}H NMR spectra (without the multiplicities) achieved by projecting the JRES spectrum (composed of chemical shift-F2 and spin-spin coupling-F1) along the F1 axis. This increases the chances of observation of better resolved peaks and therefore metabolite identification. The limitation of the method is longer acquisition times, approximately 20 min [79].

Field strength

A higher field strength of the magnet (B_0) offers improvement of signal-to-noise, but also an improvement in terms of resolution, as implied by equation 2.2 on page 21, where the signal-to-noise in an NMR experiment is enhanced as the number of nuclei in the lower energy state relative to the upper energy state increases and by equation 2.4 on page 22, hence the total signal-to-noise is proportional to (B_0 new/ B_0 old)$^{3/2}$. Therefore, using 800 MHz NMR instead of 500 MHz NMR doubles the signal. Even a modest increase in the field strength when comparing 800 MHz and 900 MHz instruments gives an advantage of almost 20 % in the signal-to-noise ratio. The frequencies of peaks are also directly proportional to the magnetic field strength. One ppm on 50 MHz NMR instrument is actually 50 Hz from resonance position of TMS, whereas on a 500 MHz NMR instrument, 1 ppm in 500 Hz from the TMS resonance position, which allows for acquisition of spectra with much better resolution.

The theory described in this section is oversimplified and a detailed introduction to NMR theory is beyond the scope of this work, however, there are many books that provide detailed discussion of the principles of NMR such as [73, 80] or more advanced [72] and [81].

2.2.2 High-resolution magic angle spinning NMR

NMR can also be used in live cells and intact tissues, often giving uniquely powerful insights [82, 83]. The method allows for non-invasive observation of different groups of metabolites and cellular processes. Similarly to liquid state NMR, this knowledge can be used in diagnostic, toxicological or environmental studies [84–87]. Tissue extracts can be analysed by high-resolution liquid state NMR

spectroscopy or mass spectrometry, however, the information on localisation of metabolites is destroyed, and targeting of groups of specific metabolites depends on the selection of solvents employed. High-resolution magic angle spinning NMR (HR-MAS NMR) spectroscopy of tissues and whole cells has shown to be valuable in assessing tissue metabolite profiles non-destructively.

In conventional liquid state NMR, molecules can tumble due to fast isotropic, or orientation independent, motion and sample all possible orientations with respect to B_0 and effectively experience the same magnetic field. In (liquid-) crystalline samples or solids, however, such motion is restricted and therefore the Larmor precession frequency becomes orientation dependent or anisotropic. Effectively, molecules present in the sample at different orientations will give different signals, where the observed spectrum is a sum of all possible orientations for each nucleus [72]. This effect, referred to as chemical shift anisotropy (CSA), gives rise to broad peaks with a characteristic shape and is presented in Figure 2.1.

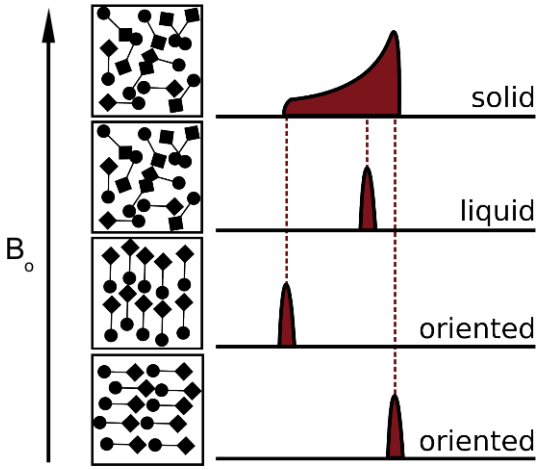


Figure 2.1: ***NMR patterns simulated for different kinds of molecular motions.*** *B_0 is the applied static magnetic field. Modified from Dr Louic Vermeer.*

In solid samples there are also sources of magnetic fields internal to the sample which may result in a number of interactions. The full NMR Hamiltonian may therefore be expressed as:

$$\hat{H} = \hat{H}_{Zeeman} + \hat{H}_J + \hat{H}_{CS} + \hat{H}_{DD} + \hat{H}_Q \tag{2.5}$$

where H_{Zeeman} is the Zeeman interaction (energy level splitting in atomic nucleus when placed in a magnetic field with the magnitude proportional to the

strength of the magnetic field), H_J is the J coupling, H_{CS} is the chemical shift coupling, H_{DD} is the dipolar coupling, and H_Q is the quadrupolar coupling.

J-coupling, also known as scalar coupling or spin-spin coupling, is a through bond coupling of nuclei. The chemical shift in NMR spectra can provide information on the structure of the molecule. Since each nucleus can be thought of as a small magnet, the orientation of that magnet has an effect on the local magnetic field experienced by other nuclei, which is mediated through bonds. An NMR peak will split into n+1 peaks, where n is the number of NMR active nuclei within 3 chemical bonds. Magnetically equivalent nuclei do not cause splitting. This situation is pictured below (Figure 2.2) using ethyl acetate as an example. The blue methyl hydrogens show a peak at 1.3 ppm. These 3 hydrogens are J-coupled to the 2 red hydrogens resulting in 3 peaks (triplet) methyl peaks (blue). The peak at 4.2 ppm for red hydrogens is split into 4 peaks, by 3 blue hydrogens (3+1).

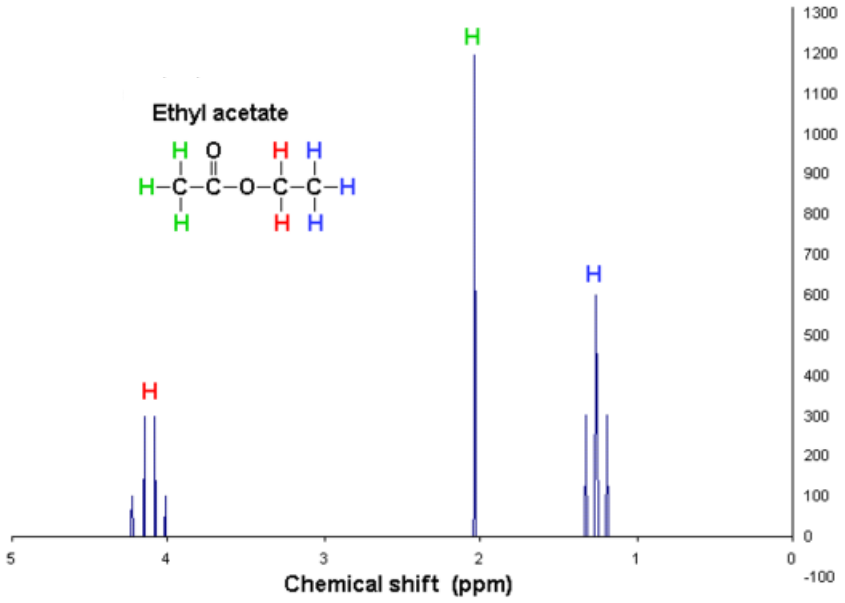


Figure 2.2: ***Simulated NMR spectrum of ethyl acetate showing the splitting pattern due to J coupling.*** *Modified from: Wikimedia Commons, the free media repository.*

The most important interaction between the spin and its surrounding spins is the dipolar interaction. It is the dominant broadening factor in organic solids. Dipolar coupling is a through-space coupling of two NMR active nuclei. It is the magnetic effect on nucleus I due to the magnetic field generated by nucleus S. The dipolar coupling of spins is either between two of the same nuclei (homonuclear

dipolar coupling) or between two different nuclei (heteronuclear dipole coupling). The natural abundance of the NMR active nuclei and the space between the nuclei directly influence the size of the dipolar coupling in solid systems. For less abundant nuclei, such as ^{13}C, the dipolar coupling can be neglected as the chance that the nuclei are close together is small. Also, as the distance between two NMR active nuclei is increased, the interaction between them diminishes. For the case of two spins, I and S, the approximate dipolar Hamiltonian can be written as:

$$H_d = \frac{1}{2}\frac{\gamma_H \gamma_C^{\hbar 2}}{r_{HC}}(1 - 3\cos^2\theta)(3I_z S_z - IS) \tag{2.6}$$

where H_d is the size of the interaction, r is the inter-nuclear distance, γ_C is the gyromagnetic ratio of the nuclei and the last term describes orientation of the inter-nuclear vector. Study of tissues and intact cells becomes possible using high resolution magic angle spinning NMR (HR-MAS NMR). The method effectively allows use of almost any solid sample, because dipolar couplings and chemical shift anisotropy are reduced according to the term:

$$(3\cos^2\theta - 1) \tag{2.7}$$

where θ is the angle between the long axis of the ellipsoid (principal tensor) with the field B_0, by spinning the sample about its own axis at the frequency of 3-5 KHz and at the magic angle (the diagonal through the cube) of 54.7° with respect to the B_0 as seen in Figure 2.3 on the next page. At this angle any vectors aligned along the z axis will be rotated through both the x and y axes, effectively making the x, y and z axes equivalent. The anisotropy is therefore removed and only the isotropic chemical shift is observed.

Andrew *et al.* [88] and Lowe [89] were first to show that spinning solid samples at the magic angle reduces line-broadening effect and results in highly resolved NMR spectra as per example from our laboratory shown in Figure 2.4 on page 29.

As with liquid state NMR, simple sample preparation protocols minimise the differences between preparations. Also, due to the rapid turnover time of some intracellular metabolites, a simple protocol without an extraction step is of great advantage for metabolomic study. In this thesis, high resolution magic angle spinning NMR (HR-MAS NMR) has been applied to bacterial cell pellet (Chapter 5 on page 107), which is a novel way of probing the bacterial metabolome.

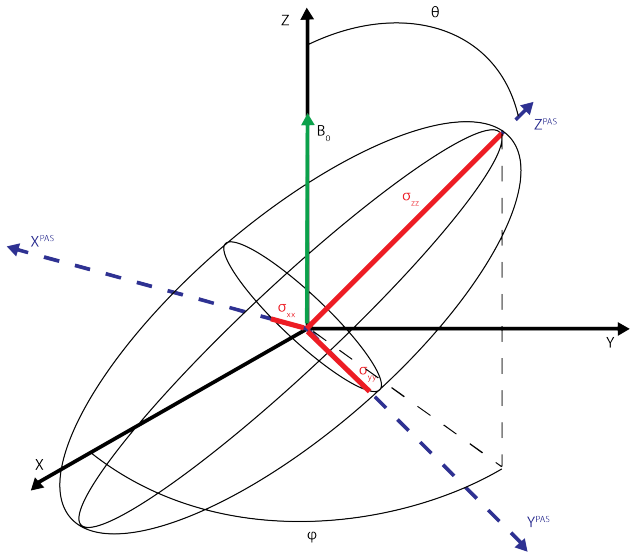


Figure 2.3: ***The distribution of the electrons about the nucleus is non-spherical; the magnitude of the shielding depends on the relative orientation of the nucleus with respect to the static field.*** *Modified from Dr Louic Vermeer.*

2.2.3 Two-dimensional NMR spectroscopy

Despite many advantages that NMR spectroscopy presents, for some samples severe spectral congestion can hinder the analysis, the identification of spectral features and assignment of the compounds. In NMR spectroscopy pulse sequences can be applied to further separate 1D NMR spectra into additional dimensions, in order to help unravel complex 1D spectra.

In two-dimensional (2D) spectroscopy, intensity is plotted as a function of two frequencies (F_1 and F_2), which map out interactions within the molecules and also between the molecules of interest. The spectrum is usually represented as a contour map with intensity peaks represented as contour lines at chosen intervals. The position of the peaks is dictated by the frequency coordinates corresponding to F_1 and F_2. Depending on the method used and the interactions probed, 2D NMR methods can be divided into three categories: through-bond coupling, through-space coupling and chemical exchange. Heteronuclear correlation experiments, such as heteronuclear single-quantum correlation spectroscopy (HSQC), are particularly useful NMR experiments where a two-dimensional spectrum is recorded in which the co-ordinate of a peak in one dimension is the chemical shift of one type of nucleus (e.g. ^{1}H) and the co-ordinate in the other dimension is the chemical shift of another nucleus (e.g. ^{13}C) which is coupled to the first

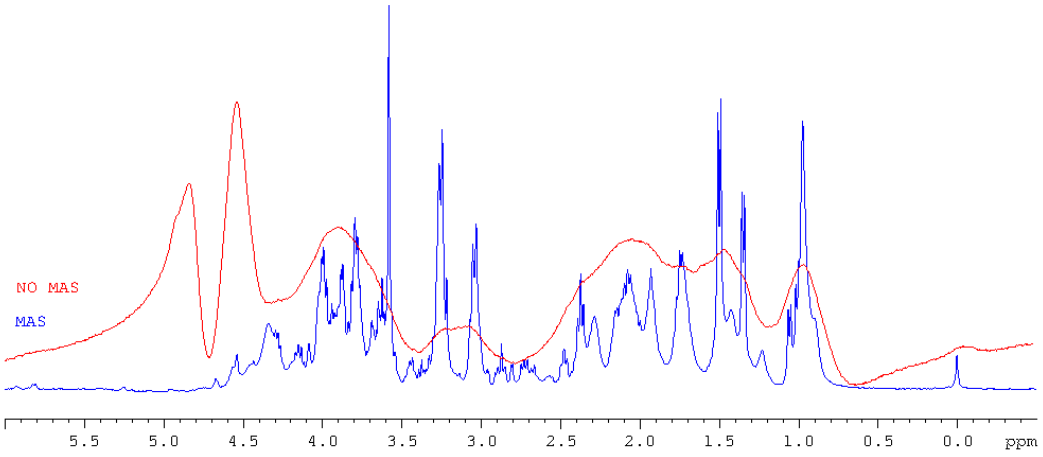


Figure 2.4: ***A comparison of a static (red) and a high speed MAS (blue)*** ^{1}***H NMR spectra*** *of human breast carcinoma sample acquired at 37°C on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm* $^{1}H/^{13}C$ *HR-MAS probe using* ^{1}H *cpmg pulse sequence.*

nucleus. Figure 2.9 on page 33 shows and example of ^{1}H/^{13}C HSQC spectrum with cross peaks for amino acid valine assigned. In homonuclear experiments, such as homonuclear correlation spectroscopy (COSY), the cross-peak indicates the position of two interacting nuclei, e.g. ^{1}H-^{1}H. In COSY cross-peaks arise due to magnetisation transfer between two nuclei that are within one to three bond lengths through J-coupling. Figure 2.10 on page 34 shows an example of a COSY spectrum with cross peaks assigned for the amino acid valine.

Two-dimensional ^{1}H J-resolved NMR spectroscopy (JRES) NMR spectroscopy has the benefits of ^{1}H 1D NMR, but has an additional dimension which allows to eliminate the congestion problem and overlapping signals, increasing the resolution and specificity of the spectral information. This method allows for visualisation of chemical shifts but also J-couplings along the second axis (Figure 2.5, C), which is instrumental for identification of metabolites in overcrowded spectra, as it provides the multiplicity and magnitude of coupling. Another advantage of JRES over other 2D techniques is that its acquisition time is much shorter (*ca* 20 min) to provide the same sensitivity as a comparable 1D spectrum, making it amendable for metabolomic study. If required an optional projection of 2D JRES spectrum along F2 can be conducted, which is similar to 1D ^{1}H NMR spectrum (Figure 2.5, A), but each proton appears as singlet irrespective if its multiplicity. Such a simplified 1D spectrum has obvious advantages in congested

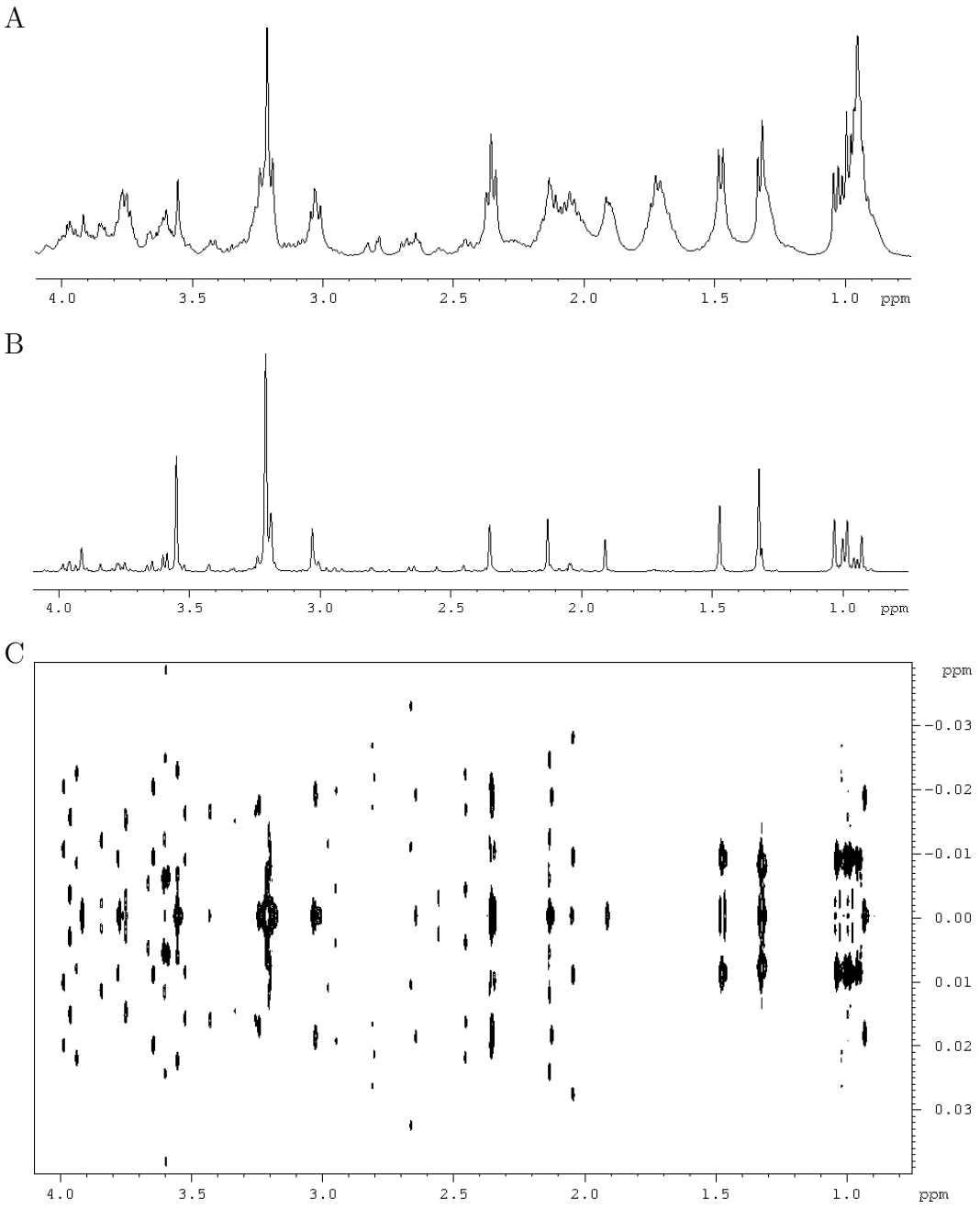


Figure 2.5: ^{1}***H NMR data of malignant human melanoma cell extracts,*** *featuring: (A) 1D NMR spectrum acquired using cpmg pulse sequence; (B) the 1D skyline projection (p-JRES) of (C) 2D JRES spectrum. All acquired at 37°C on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm $^{1}H/^{13}C$ HR-MAS probe with magic angle spinning applied at 5 kHz.*

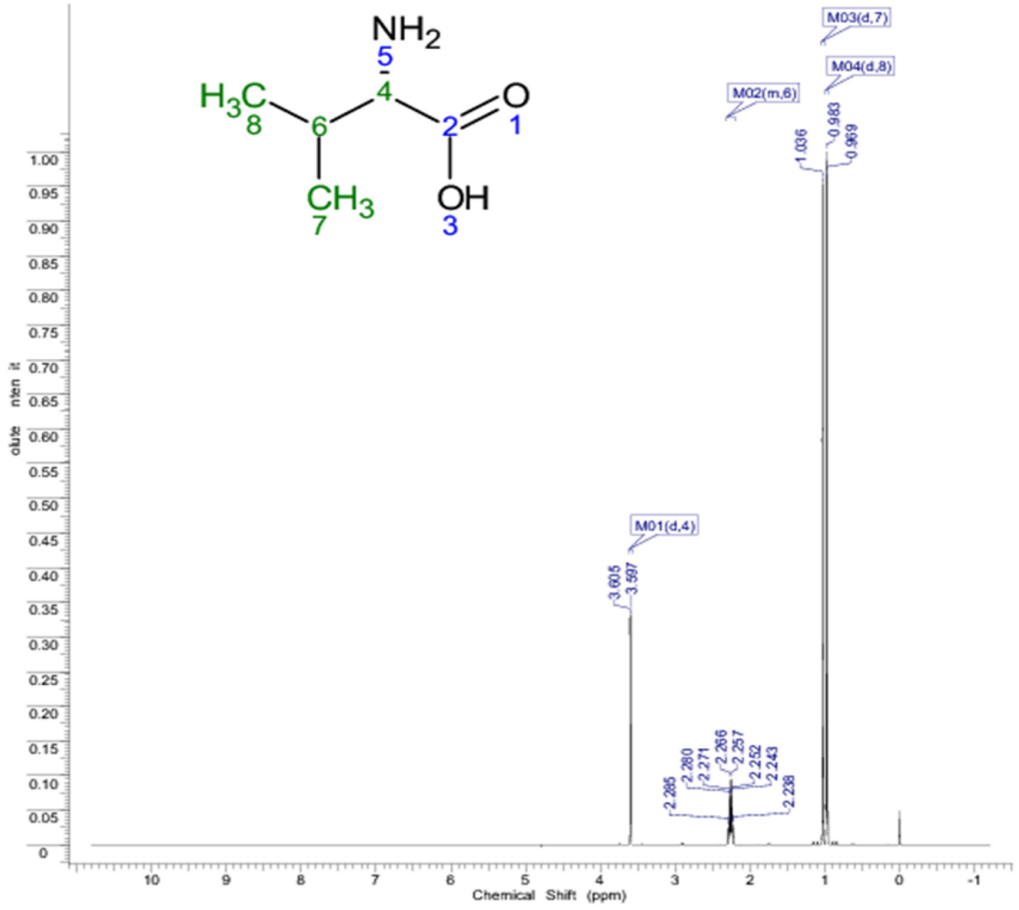


Figure 2.6: ***1D ^{1}H NMR spectrum and the chemical structure of amino acid valine*** *with atom numbers and corresponding resonances and multiplicities indicated. Taken from The Human Metabolome Database (HMDB).*

spectra [79, 90] as can be seen in Figure 2.5, B.

2.2.4 Identification

Typically, compounds are identified by comparing the observed chemical shift to those of known metabolites stored in databases such as the Human Metabolic Database (HMBD, http://www.hmdb.ca/) [91], Biological Magnetic Resonance Data Bank (BMRDB, http://www.bmrb.wisc.edu/metabolomics/query_metab.php) [92] and the *Escherichia coli* Metabolome Database (ECMDB, http://www.ecmdb.ca/) [58]. Chemical shift values, are not precise and deviations of ± 0.2 ppm or more for proton are expected. This is due to variations in the solvent and temperature in which the spectrum is being recorded [93]. Due to the nature of the metabolomic NMR spectra and the way the signals arise, the spectra of biological samples can be very complex with overlapping sig-

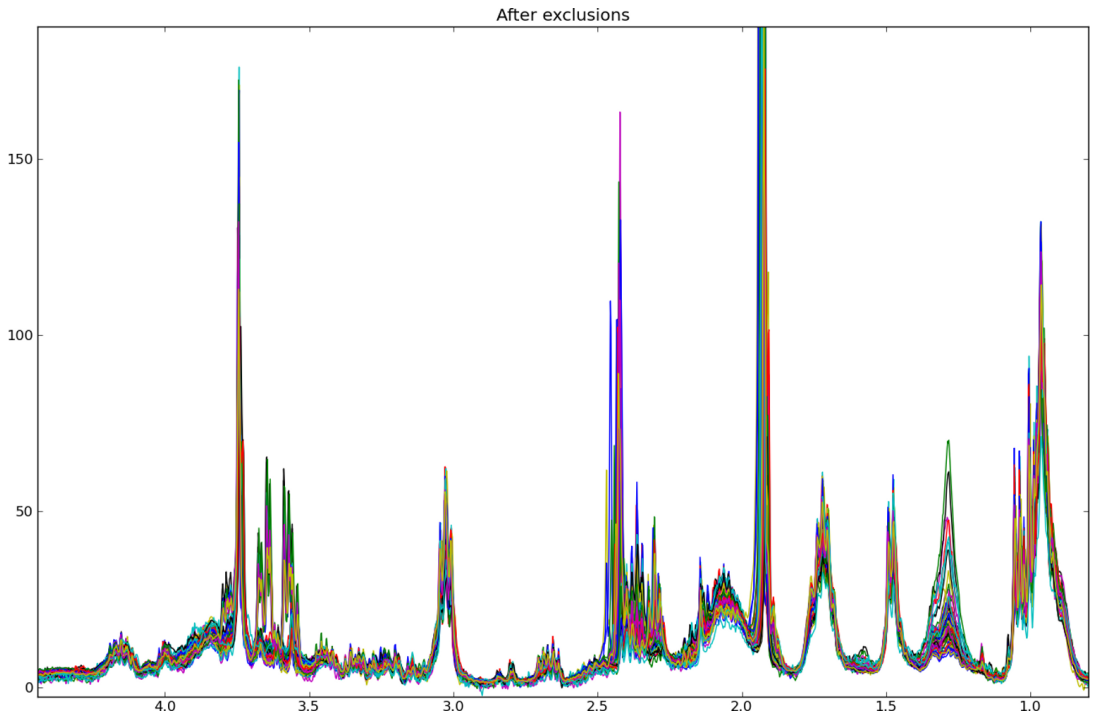


Figure 2.7: ***Overlay of the 1D ^{1}H NMR spectra*** *of* E.coli *NCTC9001 challenged with different antimicrobial peptides displaying multitude of resolved peaks. Here are shown the pellet samples acquired at 37°C on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm $^{1}H/^{13}C$ HR-MAS probe with magic angle spinning applied at 5 kHz. Different colours represent different samples.*

nals. As a result, unambiguous identification of all individual 1D NMR traces such as e.g. valine signal (Figure 2.6 on the preceding page) is challenging when looking at the recorded NMR data (Figure 2.7). Unambiguous identification of metabolites in a sample can be difficult, however the number of available databases is increasing and becoming more comprehensive. The combination of 1D spectra and multidimensional NMR techniques provides a promising framework for assignment of spectral features. Also, several commercial products are available to aid the process such as Chenomx NMR Suite (http://www.chenomx.com/software/software.php?pageID=32). Below is an example of the assignment process using amino acid valine as an example.

Some metabolites can be assigned directly by comparing the 1D NMR spectra to reference spectra from databases. This is facilitated by analysis of the signal multiplicities from 2D J-resolved spectra and experimental chemical shifts for ^{1}H and ^{13}C from 2D ^{13}C HSQC. An example of a JRES in Figure 2.8 on the following page, HSQC in Figure 2.9 on the next page and COSY in Fig-

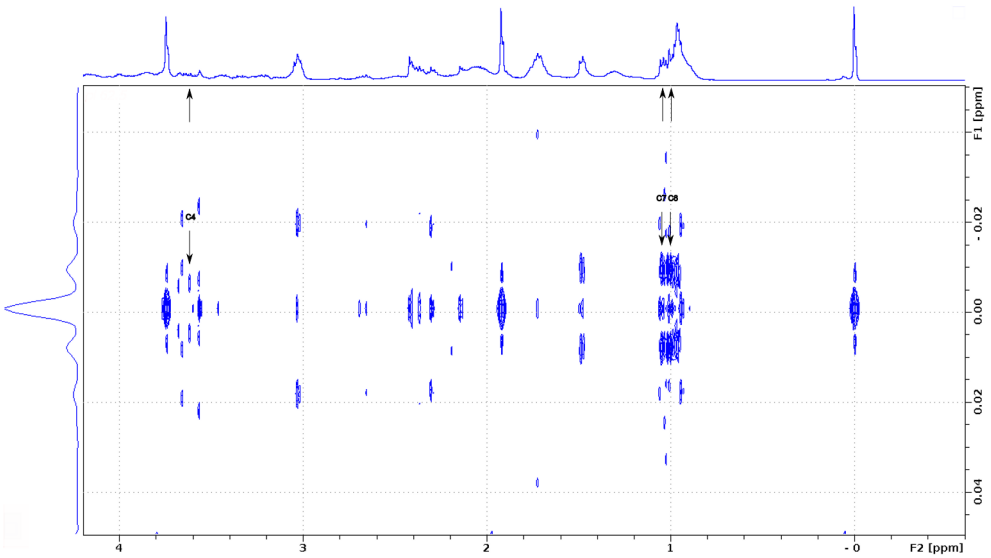


Figure 2.8: ***Representative contour plot of a 2D*** ^{1}***H J-resolved (JRES) spectrum*** *of* E.coli *NCTC9001 challenged with pleurocidin at sublethal concentration for 30 min; acquired at* $37°C$ *on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm* $^{1}H/\ ^{13}C$ *HR-MAS probe with magic angle spinning applied at 5 kHz using the standard pulse sequence. Valine peaks are identified with atom numbers annotations corresponding to those from The Human Metabolome Database (HMDB).*

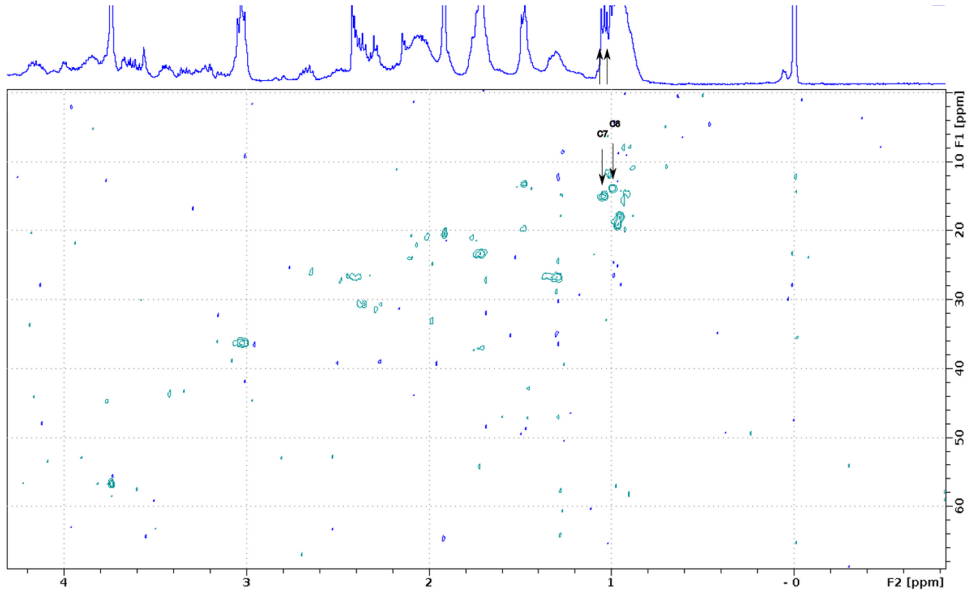


Figure 2.9: ***Representative contour*** ^{1}***H/***^{13}***C HSQC spectrum*** *of* E.coli *NCTC9001 challenged with pleurocidin at sublethal concentration for 30 min and identified valine signals; acquired at* $37°C$ *on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm* $^{1}H/\ ^{13}C$ *HR-MAS probe with magic angle spinning applied at 5 kHz using the standard pulse sequence. Valine peaks are identified with atom numbers annotations corresponding to those from The Human Metabolome Database (HMDB).*

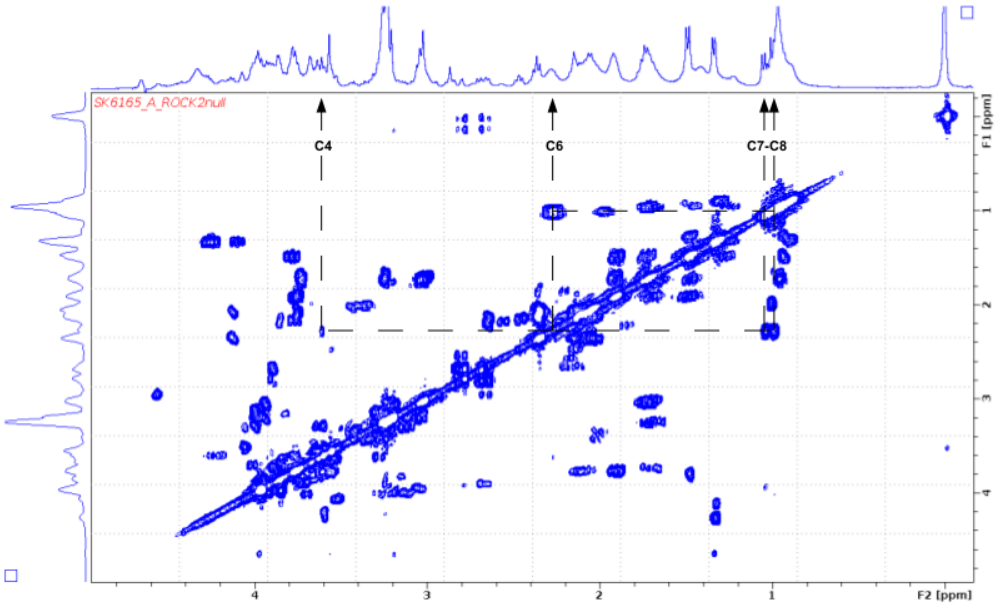


Figure 2.10: ***Representative contour COSY spectrum*** *of mouse adenocarcinoma cells and identified valine cross-peaks; acquired at 37°C on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm $^{1}H/^{13}C$ HR-MAS probe with magic angle spinning applied at 5 kHz using the standard pulse sequence. Valine peaks are identified with atom numbers annotations corresponding to those from The Human Metabolome Database (HMDB).*

ure 2.10 are shown with signals from valine identified and annotated according to atom numbers given in The Human Metabolome Database (HMDB) - Figure 2.6.

The statistical total correlation spectroscopy (STOCSY) analysis method is also used to aid identification of metabolites and confirm tentative assignments. STOCSY identifies peaks that belong to the same compound, but can also detect intra-molecular correlations and identify metabolites involved in the same pathway [94]. Figure 2.11 on the next page shows correlation between the two peaks of valine at 3.75 ppm and 0.9 ppm. The chemical shifts are in rough agreement with the chemical shifts expected from valine and provide additional confidence that the cross peaks assigned previously in HSQC, JRES and COSY belong to valine.

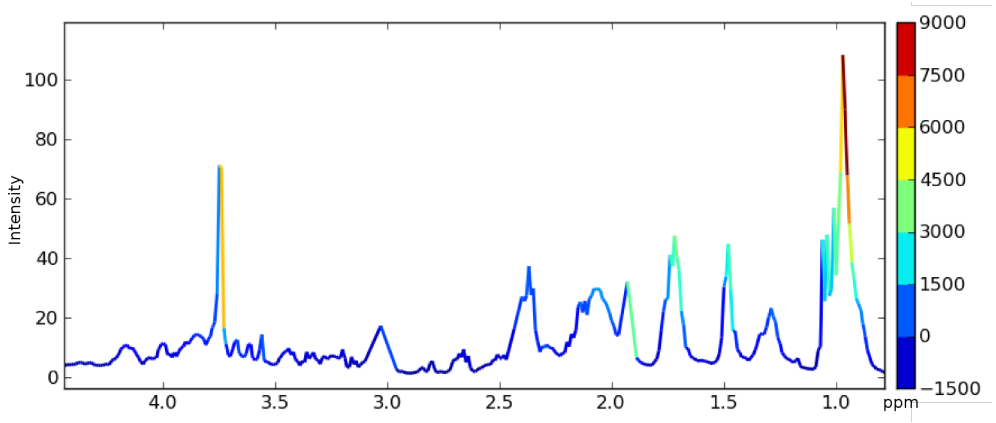


Figure 2.11: ***Representative STOCSY plot of the resonances*** *of* E.coli *NCTC9001 challenged with antimicrobial peptide pleurocidin at sub-lethal concentration from* ^{1}H *NMR spectra with two valine peaks highlighted in red, therefore showing high correlation with each other.*

2.3 Data processing and multivariate data analysis

In metabolomics, collected data sets are large, highly multivariate and challenging to interpret and derive biological meaning. Therefore various high-performance numeric computation and visualisation methods and tools are used, collectively known as chemometrics. Tools are being developed in order to tackle this challenge by various groups [95–102]. While there are many tools focused on interpretation of human metabolic experiments, the availability of options for other organisms is still limited [103]. Moreover, there is no widely accepted consensus on how to computationally process and interpret metabolic data [104–106]. The areas of disagreement include the strategy for model cross-validation, inference of class differences, number of samples necessary, parameters used to assess classification or ways to choose the overall model [107].

2.3.1 Data pre-processing

Before data exploration, spectra have to be subjected to pre-processing to make them amenable to multivariate statistical analysis. In our laboratory, software was developed using the Python programming language with numpy and scipy for calculations, and matplotlib for visualisation. The nonlinear iterative partial least squares algorithm (NIPALS) algorithm [108] was used for orthogonal PLSDA (OPLSDA) analysis (see Section 2.3.2 on page 42). A manual for data

processing using our software is included in Appendix A on page 177.

First, spectra have to be calibrated to the internal reference peak which is set to 0 ppm, phase and baseline corrected (see Appendix A on page 177), in order to avoid distortions to data analysis and also metabolite quantification. Next, NMR spectra are read into the software, aligned to a chosen peak and unwanted spectral features such as solvent peaks, residual water resonance, and reference peak are removed, by excluding the data points from the analysis. This is followed by bucketing (binning) spectra into evenly-spaced, user defined segments usually of 0.01-0.04 ppm size, in order to correct any minor, global peak shifts (as opposed to local), decrease the noise and reduce the number of data points to increase the speed. The intensities inside each bin are summed, so that the area under each spectral region is used instead of individual intensities. The downside of this approach is that it may reduce spectral resolution, while some peaks may appear in two or more bins, splitting the chemical information due to a lack of flexibility of the boundaries. This can be somewhat overcome by applying manual bucketing to problematic spectral regions that are not crowded. Bucketing is, nevertheless, a commonly applied pre-processing procedure. Various binning algorithms have been proposed recently [109–112], however, their applicability has not been tested in our laboratory.

Sometimes complex spectra cannot be aligned by the binning method or the peak shifts are larger than the bin size. This can be due to instrumental instabilities, temperature variations, or variations in pH and ionic strength, both of which influence the ionisation state of basic or acidic groups and thus their associated chemical shifts. It has been demonstrated that peak shifts can be beneficial to discern between different groups of samples under certain circumstances, however, this effect is unwanted for most applications [113]. Shifting peaks have to be corrected or excluded from the analysis as they can hamper the discovery of patterns in the data and lead to incorrect interpretation of the results. Peak alignment methods are more elaborate than bucketing and generally require use of commercially available software. Giskeodegard *et al.* [114] reviewed the five most prominent warping methods: interval correlation shifting (icoshift) [115], correlation optimised warping (COW) [116], fastpa [117], variable penalty dynamic time warping (VPdtw) [118] and parametric time warping (PTW) [119].

In our laboratory a freely available COW-based method adopted for use with

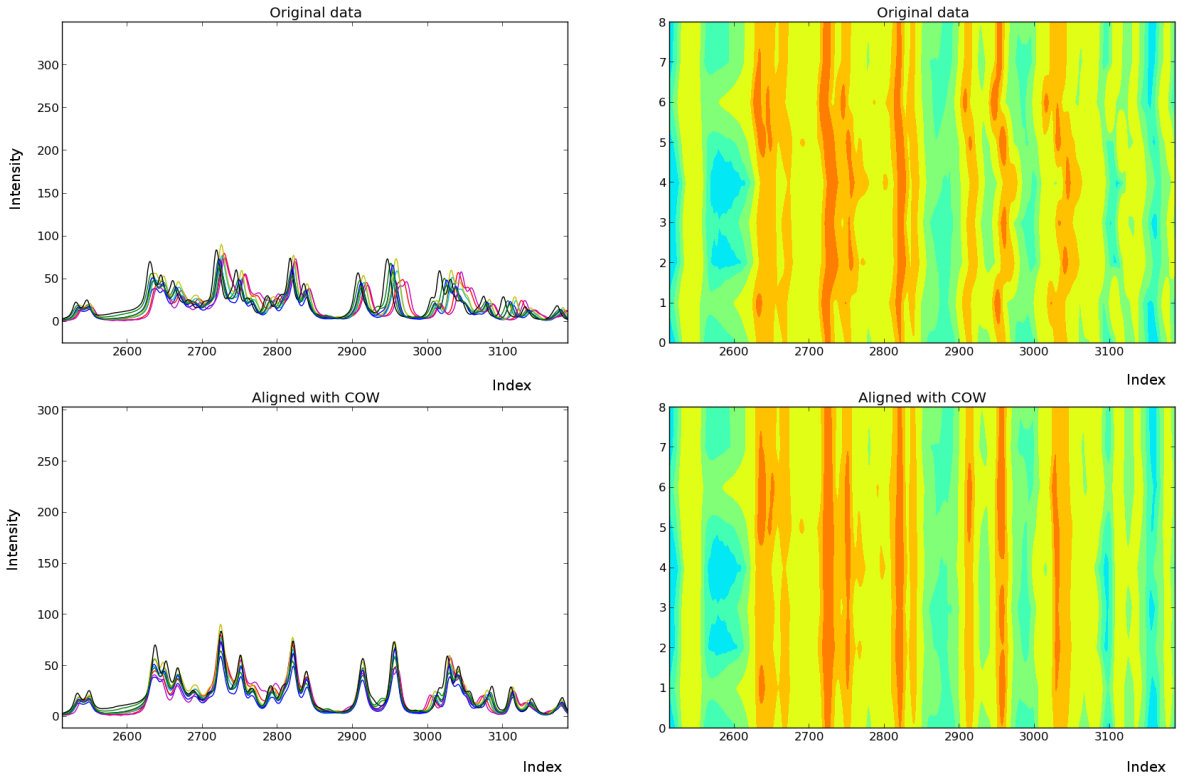


Figure 2.12: ***An example of*** ^{1}***H NMR spectra of*** **E.coli** ***NCTC9001 challenged with pleurocidin at sublethal concentration before (top panel) and after (bottom panel) alignment using COW****; acquired on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm* $^{1}H/^{13}C$ *HR-MAS probe with magic angle spinning applied at 5 kHz using the standard cpmg pulse sequence.*

Python [120] is used (see Appendix A on page 177), based on the original paper by Niels-Peter Vest Nielsen *et al.* [116]. Previous packages used in this project have been freely available and we wanted to maintain this and the software also allows for adjustment of many parameters in order to achieve the best results. It uses two input parameters based on the width of the peaks-m, which is the segment length and t, which is the amount of slack, i.e. the maximum range or degree of warping in segment length. It has been suggested that parameter m should be at least equal to the width of the smallest peak that has to be aligned and that lower values may result in the alignment of noise and distortions in peak shapes [116]. For each set of spectra, optimal combination of these parameters has to be established. Figure 2.12 shows an example of spectra aligned using this method. Notably, spectral warping can distort spectra and introduce artefacts, thus absolute quantification should be performed on unaligned spectra.

This is followed by normalisation, in which spectra are scaled to the same overall concentration to account for differences in amount of material per sample, overall variations in sample concentration or for technical reasons in order to make the samples directly comparable and amendable to multivariate analysis. Generally, normalisation involves multiplication of every row (i.e. sample) by a constant specific to each sample. In the total integral normalisation method the constant is the total integrated intensity across the whole spectrum or part of it so that each data point is expressed as a fraction of the total spectral integral. Total integral normalisation is the standard method used in most metabolomic studies [121–123], however, it is not optimal for spectra with extreme concentrations of certain metabolite in the sample, as this will skew the total integral and hamper the subsequent scaling and analysis. This issue is overcome by the probabilistic quotient normalisation (PQN) [124] method, which is currently the method of choice in our laboratory. Here the spectra are normalised using the most probable quotient. The most probable quotient can be calculated from the distribution of signals in a spectrum divided by the reference signal (e.g. median or mean spectrum from the study or a reference spectrum) [124] so that it is not affected by the peaks with large changes.

The last step in data pre-processing is scaling. Metabolite intensities have to be scaled to adjust for variations between high- and low-concentrated metabolites in a sample. The more abundant metabolite will give rise to higher peak, which will be considered more significant in data analysis, as higher intensity is generally linked to higher variation. A number of scaling methods are in use. Mean-centering moves the centre of the data to around zero, instead of the mean of each intensity, calculates the average spectrum of the data set and subtracts this from each spectrum. Mean-centering is usually insufficient, particularly for data of bivariate/multivariate distribution and it is often used in combination with other scaling methods, as it moves the origin of all the components in principal component analysis (PCA) (see Section 2.3.2 on page 40) to the centroid of the data giving a parsimonious model. Another widely used form of scaling is autoscaling, where mean-centering is followed by division of each column (intensity) by the standard deviation of that column. In some cases, autoscaling is applied in order to amplify signals of low intensity which can carry important information but are obscured by high intensity peaks. This process, however, also amplifies the noise and therefore should be used with caution and the results always checked against the original spectra. Pareto scaling is considered as a

compromise between mean-centering and autoscaling. In the process of Pareto scaling, data are mean-centered and divided by the square root of the standard deviation. In this work, the input variables are preprocessed by autoscaling with a few exceptions where Pareto scaling is used (with justification).

There is no optimum pre-processing method and a combination of processes described above has to be applied depending on the experiment, sample origin and analytical method used for data acquisition. The procedures described above highlight the importance of a good sample preparation protocol due to the nature of the NMR spectroscopy (low sensitivity, crowding of spectra, line broadening due to viscosity of the sample) and also the need for spectral reproducibility, which is essential for metabolomic study.

2.3.2 Model construction and data classification

Despite considerable progress made in the field of metabolomics, the extremely large and highly correlated datasets still remain a challenge to analyse and interpret. Typically, in a metabolomic dataset there are more variables than samples (100-100,000s) and the data are highly correlated and non-normally distributed. Unlike in e.g. microarray data, the number and identity of compounds is not known. For these reasons different multivariate data analysis tools have been developed. They enable extraction of information from the data in order to visualise the trends and relationships between the samples and variables and subsequent generation of predictive models. This can determine which molecular entities are 'associated' with specific disease, perturbation, condition etc. It is also possible to determine which subset of thousands of chemically diverse compounds present within a cell provides a unique signature for a specific disease, perturbation, condition etc. Knowledge of how those chemical entities are functionally related and how they respond to perturbations allows to map them onto known metabolic pathways which can provide more comprehensive, pathway-based interpretation.

Analysis of metabolomics data involves first spectral analysis of the data showing the overall changes in metabolite concentration. At this level, the study is neither qualitative, nor quantitative, but should be thought of as a metabolic fingerprint reflecting the response of the organism to the changing environment [11]. The data is then subjected to statistical analysis and classification models are generated. For the results of such analyses to be meaningful, they must be reproducible over a period of time. Very often, reproducibility is degraded due to

analytical and sample differences introduced during metabolite extraction procedure [125] therefore the simplified and straightforward HR-MAS NMR sample preparation has an obvious advantage. Next, the changes in the metabolome are identified and quantified. This involves dividing the datasets into classes according to the level of variation they demonstrate and determining how well the data fit with the predetermined classes using statistical software and identifying the compounds that differ between classes. The final step is to assign them to specific pathways, which provides information about the intrinsic mechanisms [11]. Principal component analysis (PCA), partial least squares regression or projection to latent structures (PLS) and PLS discriminant analysis (PLSDA) are, by and large, the most commonly used multivariate analysis methods that can be applied to metabolomic data.

Principal component analysis (PCA)

Principal component analysis (PCA) was first introduced in 1901 by Karl Pearson [126] and later developed and named by Hotelling [127], while a good modern reference is Jolliffe [128]. PCA is a method used to reduce the dimensionality of multivariate data, whilst preserving as much of the relevant information as possible, and also to gain an initial insight into hidden patterns and relationships in acquired data. During the PCA analysis the original measured variables are transformed into new variables referred to as principal components (PCs), where the greatest variance is explained by the first PC, the second greatest variance by the second PC, and so on. Usually only two or three principle components are sufficient to explain most of the information in the collected data. The data can be plotted in a coordinate system based on two or three largest principle components, that are always uncorrelated and orthogonal to each other. That way each PC represents different type of independent information since it defines a different direction of variance. PCA is used mainly to discern information about the overall structure of the data i.e. similarities and differences among samples. It is also used for identification of outliers, usually spectra of unacceptable quality that can be potentially excluded from the analysis. PCA is an unsupervised method, as the analysis is blind i.e. it does not require any prior knowledge, and it relies entirely on the input data itself. This ensures an unbiased examination of the data.

Figure 2.13 on the following page (top right panel) presents the scores plot of PC1 and PC2 for a model with 3 classes. Each marker on the plot represents one

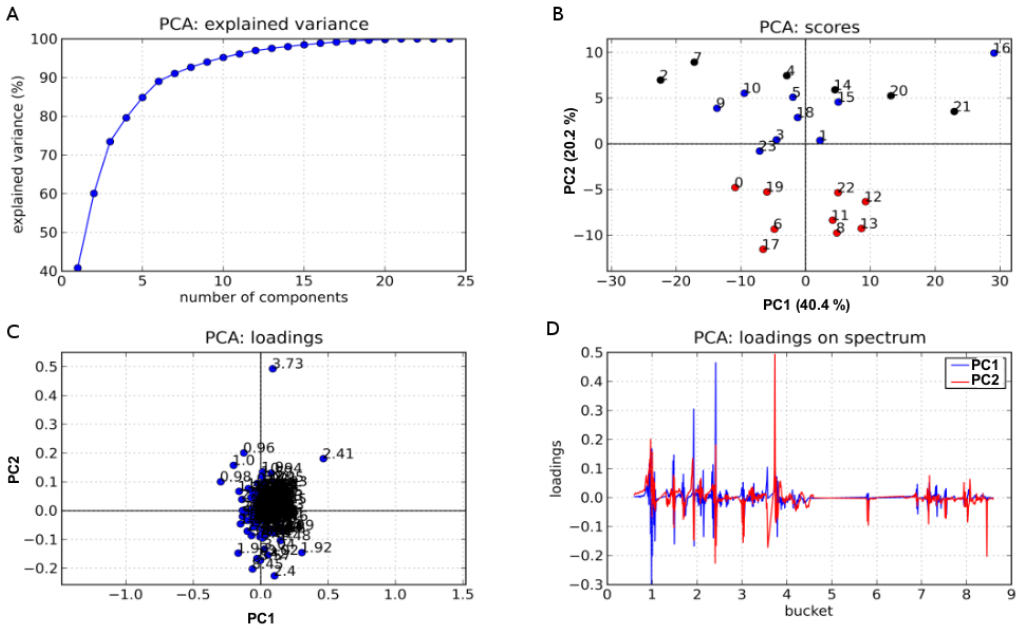


Figure 2.13: ***A representative PCA analysis*** *of samples containing 3 classes: class 1 (black), class 2 (red), class 3 (blue). PC1 and PC2 account for 60.6 % of the total variance in the data (A). PCA scores indicate the clustering pattern (B), whereas loadings indicate the features of the NMR spectrum contributing to the separation seen in the scores plot (C/D).*

single sample. In this example, spectrum 16 is a possible outlier, as it clusters away from other blue markers. Samples from the class 1 and 3 cluster together in the top quadrants, therefore they influence the model in a similar way and there is not much variation between samples of those two classes. Samples belonging to class 2 (red markers) are separated from class 1 and 3 (black and blue markers along PC2 axis. The further away the sample from the origin, the more dissimilar the sample is from samples belonging to class 1 and 3 and therefore the more influential the sample is on the model. The loadings plot (Figure 2.13 bottom left panel) describes the importance of the variables, which are the peaks in the NMR spectrum. The direction in the loading plot correspond to those on a scores plot and the loadings are the ppm values from the NMR spectrum. The loadings plot, therefore allows interpretation of the score plot by examination of the loadings and correlates the values to the original NMR spectra. In this example the peak at 3.73 ppm is the main spectral feature separating classes 1 and 3 from class 2 along PC2.

Partial least squares (PLS) and orthogonal projections to latent structures (OPLS)

PLS (also known as projection to latent structure) is the most widely used supervised statistical method and has been used in science and technology since 1980 [129]. The technique models the variables (**X** block) using a set of predictor variables (**Y** block), which are given by the user. PLS finds the components (latent variables) which discriminate between two or more different groups of samples using their covariance with the predictor variables (target class, e.g. control vs treatment). This method is used for regression modelling when data represent continuous variables. When variables are in a discrete form and represented as class memberships, PLS discriminant analysis (PLS-DA) is applied. This method is used in classification and biomarker studies. In PLS-DA the results are presented in the form of a scores plot, which gives information about class membership and separation. The weights plot provides information about which peaks in NMR spectra are responsible for the separation. PLS-DA is a latent variable method, therefore the data can be projected onto the new space so that it maximises the covariance between scores in **X** and **Y** spaces, unlike in PCA, where the algorithm is just looking at the variance in the data set. It is used for two-class models. Orthogonal PLS (OPLS) is a modified PLS method, which divides variation in the data into correlated and orthogonal components [107, 130]. The predictive components are linearly related to the response, whereas orthogonal components contain unrelated information, such as technical issues or unintended differences in experimental conditions. The models generated using OPLS are the same, as using PLS, and therefore the predictive power of the models is identical, however, model visualisation and subsequent interpretation of the data is improved [131]. The method can be extended to OPLS-DA.

Hierarchical cluster analysis and generation of heatmaps

Hierarchical cluster analysis (HCA) is another unsupervised method similar to PCA, which is routinely used to show similarities within metabolic data [132]. The samples are grouped pairwise based on the similarities between them. The results are presented as a dendrogram where the length of the branch corresponds to the difference between samples and their groupings. The algorithm constructs a hierarchy from top to bottom on the basis of a self-organising tree. It dynamically finds the number of clusters at each level. Figure 2.14 on the following page shows a representative heatmap from HCA of changes in the metabolite levels in

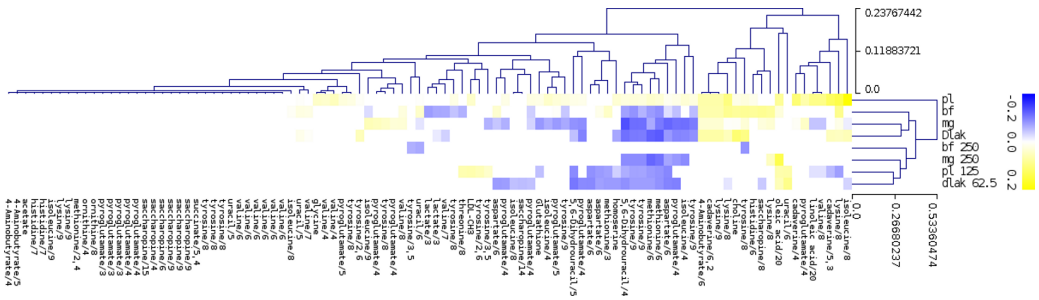


Figure 2.14: ***Representiative clustered heatmap of the metabolite yields*** *(columns) in* E.coli *NCTC 9001 challenged with different AMPs (rows).*

bacteria in response to treatment with different AMPs. Different treatments are presented in columns and in rows are metabolites that differ in yield depending on the treatment. HCA groups metabolites based on significant differences in relative abundance. The level and scale of clustering has to be chosen depending on the application. Since cross-validated data is used, a Euclidean distance algorithm was chosen to calculate the distance between every pair of objects in a data set, since it computes the differences directly from the data already subjected to the statistical analysis in the most intuitive way. Either the average linkage or complete linkage method was used to link objects into binary clusters based on distance information from the metric algorithm. Complete linkage computes the distance between two clusters as the distance between the two farthest objects in the two clusters, whereas average linkage computes the distance between two clusters as the average distance between objects from the first cluster and objects from the second cluster. The freely available MultiExperiment Viewer (MeV), which is a part of the TM4 Microarray Software Suite [133], was used for hierarchical clustering analysis and generation of heatmaps. The details can be found in Appendix A on page 177.

2.3.3 Model validation: OPLS cross-validation

A vital stage of classification analysis is model validation, also referred to as cross-validation. It is a standard resampling procedure which estimates the performance of a model when applied to unknown data and checks if it does not overfit the data. It allows the overall complexity of the model i.e. the number of PCs in a PCA model to be assessed [134]. Generally, cross-validation procedures involve a series of iterations, where, in each iteration, a subset of objects from the dataset is removed (the test set) and a model is constructed using the remaining objects in the dataset (the model building set or training set) and

subsequently the created model is used to predict the removed objects. Typical cross-validation involves many iterations and in each round a different subset of samples is chosen as a test set and training set [134]. Models are therefore tested using a full training set by means of repeated resampling in a systematic manner. Maximising, in that way, the number of data points used for testing helps to protect against overfitting of the data. There are different cross-validation techniques which will depend mainly on the method of sample subset selection for each iteration [134]. The choice of cross-validation method is dictated by the data: the number of samples in the training and test sets, the inherent correlation in the data and ordering of the samples and the total number of variables or the presence of replicate samples. Usually the method of choice is the one generating the lowest predicted residual sum of squares (PRESS). In our laboratory, data are subjected to leave-one-out cross-validation (LOOCV) [135, 136], where part of the samples are used as a training set and the remaining samples as the test set, ensuring that the number of samples in the test set is proportional to the total number of samples from each class and that at least one sample from each class is present in the test set. The selection of the optimum number of components is performed using the lowest prediction error in cross-validation carried out on the samples in the training set, i.e. the optimal prediction of the training set using test set (samples excluded in the calibration step) or the highest F-1 score selection method for the two-class models. This method finds balance between false positives (class 1 predicted but actually class 0) and false negatives (class 0 predicted but actually class 1), as simply counting the number of correctly predicted samples would be unfair when the number of samples between classes is not equal. This double cross-validation is repeated 2000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. The size of both training and test set affects the classification result. With small sample size only the major differences between groups will be detected, whereas larger sample size allows for more robust model and more representative results.

OPLS cross validation provides a scores plot, which shows whether there is any difference between the two classes of samples and in this case dots cluster away from each other indicating that there are features in the dataset, that allowed discrimination between the two classes. The backscaled loadings plots present the OPLS model that was 'back-scaled' by plotting the variables with the colour scale representing their respective correlation weights. This results in a loadings

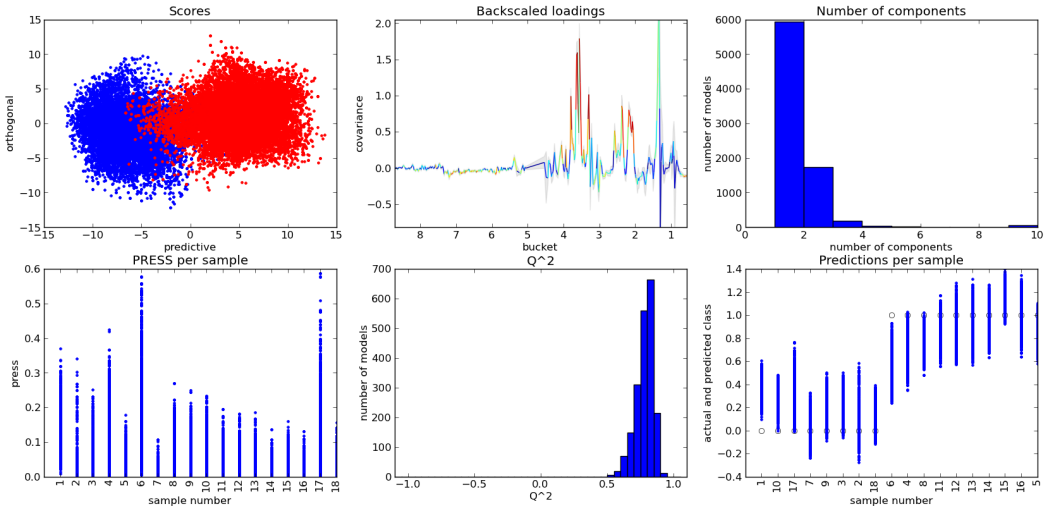
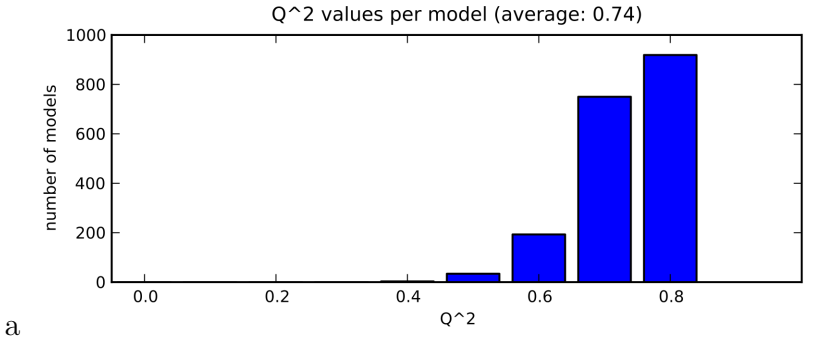
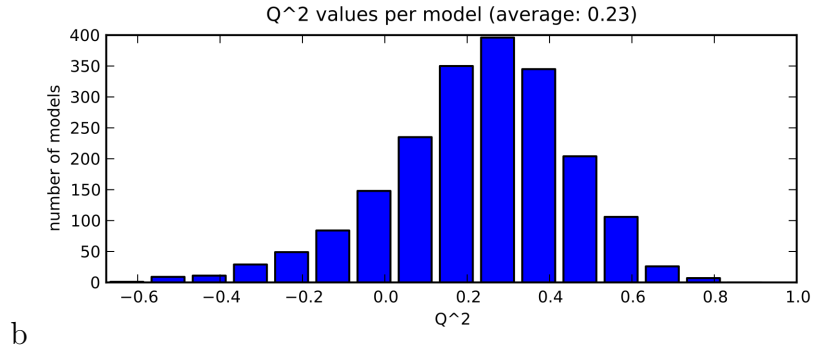


Figure 2.15: ***Representative cross-validated metabolomic analysis by ^{1}H HR-MAS NMR.*** *In scores plot blue and red dots represent scores from two different treatments. Plots discussed in the text.*

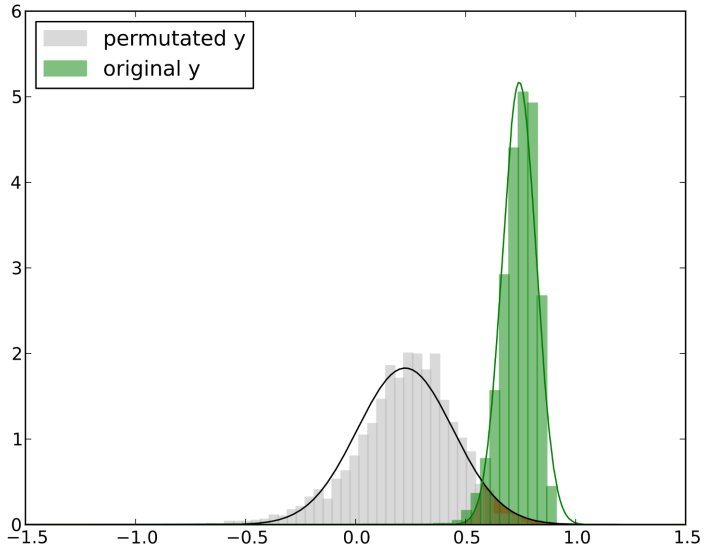
coefficient plot [113]. The main advantage of the plot is that it shows the variables with the discriminatory importance directly on the spectrum. The variables that are important for discrimination between classes are mapped in red and orange, whereas variables with negative correlation are mapped in blue. The next panel in the plot in Figure 2.15 indicates that only 2 components were sufficient to discriminate between the two classes of samples with low PRESS per sample (the following panel). Histograms of the number of times that a particular Q^2 (goodness of fit) value was obtained in each of the rounds of resampling is shown on the next plot. The Q^2 value was calculated as $Q^2 = 1$ - (PRESS/TSS) where PRESS is the sum of squared differences between the known and predicted classes, and TSS is the sum of squared differences between the known classes and their average (= the total variance). A Q^2 of 1 represents the perfect score, whereas a score below 0 indicates no discriminatory power of the model and it is generally considered to be good when its value is higher than 0.5 [107, 137]. Finally, this procedure is repeated and the results are compared to a reference values obtained by computing Q^2 for models where the classes were assigned randomly, i.e. the cross-validation procedure was repeated again after randomising the $\mathbf{Y}$-table (the classifiers) [107, 137]. It allows to measure the performance and stability of the models and is referred to as permutation tests. By comparing obtained Q^2 of the model (Figure 2.16 on the next page, a) with the Q^2 of the randomised model (Figure 2.16 on the following page, b) one can see if the model performs well. The random data should give lower Q^2 than the data set and in that way one knows



a



b



c

Figure 2.16: ***Visual evaluation of the permutation test for the significance testing.*** *a) Q^2 distribution for the D-LAK threshold model and b) Q^2 distribution for the random class assignments, c)The red-shaded area of overlap between two histograms.*

which Q^2 value corresponds to a good discrimination between groups [107].

2.3.4 Summary

This chapter provides an overview of tools and methods used in metabolomics. It specifically explains the applications of NMR and solid-state NMR in the context of metabolomics and presents the steps involved in the pre-processing and processing of the data. The chapter also provides an overview of the multivariate data analysis methods used in the field and explains the rationale behind applying each of them.

Some of the key issues that researchers in the field of metabolomics are faced with are also covered. Metabolomic data is inherently complex and biological contextualisation difficult and requiring knowledge of metabolic networks and tools. This is compounded by the fact that there is no single analytical method that can detect all the metabolites within the system due to chemical heterogeneity and there are no standardised methods to process and interpret the data.

Some of those issues will be solved to an extent by technological advances and development of new methodologies, while efforts in development and curation of community databases will help metabolite identification. The final steps of analysis, biological interpretation, have considerable potential for improvement and the field will benefit from emerging pathway enrichment and visualisation tools [63].

3

Metabolomic investigation of a relationship between *Pseudomonal* growth behaviour and cystic fibrosis patient lung function

Work described in this chapter has been published as:

Kozlowska, J., Rivett, D.W., Vermeer, L.S., Carroll, M.P., Bruce, K.D., Mason, A.J. & Rogers, G.B., A relationship between *Pseudomonal* growth behaviour and cystic fibrosis patient lung function identified in a metabolomic investigation. *Metabolomics* 2013 (9) 1262-1273

3.1 Introduction

Cystic fibrosis (CF) is the most common autosomal recessive inherited disorder in the Western world [138] and in UK almost 10,000 people are affected with more than two million people being carriers of the faulty cystic fibrosis transmembrane conductance regulator (CFTR) gene [139]. CF is caused by loss-of-function mutations of the CFTR gene, that results in production of abnormally viscous secretions by epithelial cells lining the surfaces of lungs, pancreas, liver, reproductive track and intestine of patients. Almost 2000 gene mutations have been described in CFTR, which can be grouped according to their disruptive mechanism on CFTR function, and association with residual function and disease severity [140]. While CF is a multi system disease, the primary cause of death in patients with CF is respiratory failure [139]. CF is a progressive condition and over time infection becomes established and chronic by adulthood. Chronic lung infections in CF patients are typically dominated by high levels of *Pseudomonas aeruginosa* (*P. aeruginosa*) and its presence is associated with reduced life expectancy [141]. *P. aeruginosa* was cultured in specimens from 61 % of all patients, ranging from 21 % of those less than 1 year of age to more than 80 % of those aged 26 years or older [142]. *P. aeruginosa* is an opportunistic pathogen and its virulence stems from a multitude of factors; it is able to form biofilms, secrete toxins and is intrinsically resistant to various antibiotics as well as host defences [141]. Periodic exacerbations (episodes of sub-acute worsening) of *P. aeruginosa* respiratory infection in patients with CF have traditionally been treated with antipseudomonal antibiotics. Despite frequent intravenous therapy, patients continue to have a decline in pulmonary function of approximately 2 % per year, and eventually 90 % of such patients die of lung disease [142]. There is still no cure for CF, however, more than 50 % of the CF population in the UK will now live beyond the age of 41 thanks to many treatments, including antibiotics, physiotherapy, exercise and nutrition, available to manage the condition. This is a significant improvement since 1999, when only half of sufferers would live beyond 30 years [139].

Undoubtedly, improved use of antibiotics is responsible for a substantial portion of the increased survival that has occurred in patients with CF and antibiotic therapy remains essential component in the management of CF lung disease. Three distinct antibiotic strategies are used depending on the stage of the infection. First, early aggressive antibiotic therapy against *Pseudomonal* infection has

been advocated to delay onset of chronic *P. aeruginosa* colonisation [143, 144]. Once colonisation with pathogens such as *Staphylococcus aureus* (*S. aureus*) and *P. aeruginosa* is established chronic maintenance antibiotics are prescribed to minimise decline in lung function and reduce the frequency and severity of exacerbations of pulmonary symptoms. During exacerbation intensive antibiotic regimens are frequently administered to relieve symptoms and restore pulmonary function to baseline values [143, 145]. The choice of appropriate antimicrobial therapy should be based on review of recent cultures of airway secretions, but a combination of an aminoglycoside and beta-lactam is recommended to provide synergy and slow emergence of resistance [146]. Antimicrobial resistance is a major problem in CF patients and increasing numbers of patients with chronic *P. aeruginosa* infection develop multiresistant strains resistant to all drugs in at least major classes of antipseudomonal antibiotics: beta-lactams, aminoglycosides, and quinolones [145].

3.1.1 Metabolomics for diagnostics of CF

It has long been known that bacteria can influence their environment by secreting a large number of metabolites. *P. aeruginosa* growth influences metabolite production, consumption, and biotransformation and therefore has the potential to play a selective role in microbiota composition in the mucus in the CF lung influencing the course of the disease and severity of the symptoms. Therefore establishing the impact of *P. aeruginosa* growth on airway secretion composition is fundamental to understanding the behaviour of this pathogen *in vivo*, its impact on the host and relationship with other colonising species.

While high-throughput sequencing and microarrays have proven successful to interrogate 16S rRNA gene sequences in defining the components of the microbial community in the CF lung [147], a metabolomics approach was proposed to determine the functional impact of isolates on the CF lung. The CF biofilm is not a sum of its bacterial components, but a complex polymicrobial community and the snapshot of the dynamic state of the CF lung can be obtained using metabolomics [52]. Such variation in composition also poses technical problems to current diagnostics which are culture-based and often unable to isolate all potentially clinically significant bacterial species present in a sample. In this study a systems biology approach will be applied utilising the research platform based on NMR metabolomics. It was hypothesised that the phenotypes of clinical isolates would be reflected in the differences in the metabolite levels in spent media

and that those changes might correlate with patient outcome. Since the bacterial species in CF vary markedly between patients, being able to relate the microbiota composition, particularly the dominant *P. aeruginosa* strain characteristics, to the lung function in a quick and high-throughput manner would be highly beneficial. It would allow better diagnosis, treatment and, in the future could possibly inform new strategies to tailor the composition of bacterial community in order to promote or silence certain bacterial behaviour *in vivo* and predict the effects of new therapeutic interventions. Identification of the key biochemical signatures of *P. aeruginosa* at different time points of the infection and at varying virulence of the pathogen would allow more effective treatment and prophylaxis; the limitations of currently used diagnostic methods are well documented [148]. Of particular importance is prevention of phenotypic diversification of the pathogen and formation of biofilm by aggressive antibiotic prophylaxis [149].

3.1.2 Development of CF airway composition

The observation that the variation amongst species colonising CF airways is much lower than in the airways of a healthy individual [147, 150] and also the presence of a consistent microbiome pattern of the core species [151] suggests involvement of a selective mechanism in the infection. This is supported by the relative stability of the overall composition in the adult patient lung during antibiotic treatment and exacerbations [152], which suggest that the colonising species have successfully adapted to the environment and that they are able to selectively compete or cooperate with other species.

Van der Gast *et al.* identified clinical factors that can influence the composition of CF lung, which are CFTR genotype and recent antibiotic treatment [151]. Interestingly, factors such as patient age, gender or FEV_1 showed no significant correlation with the composition of CF lung microbiota [151], however, the study ignores the input from the dominant strain *P. aeruginosa*. Nevertheless, this finding emphasised the need for a new method of assessing CF patients outcome, as forced expiratory volume in 1 s (FEV_1) is currently used as the single main estimator of mortality and is also used as a main classifier for lung transplantation [153], despite the fact that it does not correlate with the lung composition and dynamics. Interestingly, preliminary data analysis of the samples showed strong negative correlation between mean FEV_1 and spent culture pH ($R = -0.76$, $p = 0.002$) (Figure 3.1 on the next page). No significant correlation was found

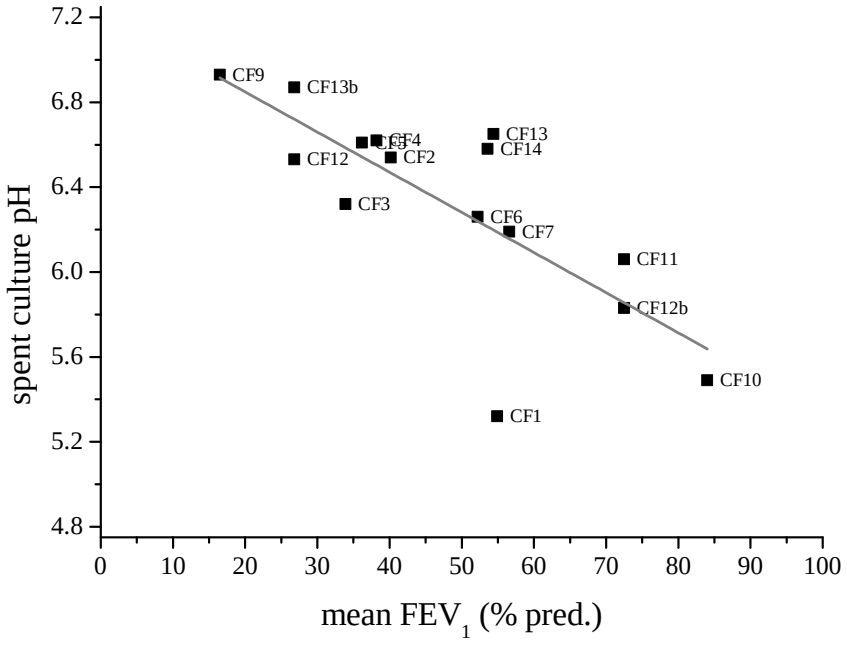


Figure 3.1: ***The relationship between mean FEV$_1$ and spent culture pH shown for each of the isolates*** *(R = -0.76, p = 0.002). Analysis performed by Dr Damian Rivett.*

however between lung function and sputum pH (R = 0.50, p = 0.067) or between sputum pH and spent culture pH (R = -0.37, p = 0.188). Tables 3.2 on page 56 and 3.3 on page 57 list further significant correlations. Therefore a further objective was to investigate if culture pH also correlates with the nutritional changes.

Secretory products of pathogens are known to contribute to the survival strategy. Alkaline protease secreted by *P. aeruginosa* is involved in the activation of the epithelial sodium channel (ENaC), leading to further impairment of mucociliary clearance [154]. Another pathogen, *Burkholderia cenocepacia* (*B. cenocepacia*), secretes lipase, which promotes its epithelial invasion, whereas alginate production by *P. aeruginosa* mucoid phenotype enhances persistence of *B. cenocepacia* [155]. Many bacterial pathogens are also known to secrete toxins targeting their competitors residing in the same niche. Here, the differences in the metabolic composition of the secretions will be investigated. Colonising bacteria can influence the emergent properties of the community and this in turn can be a modulating factor [24]. Since *P. aeruginosa* is the dominant isolate, its growth will be the major factor affecting availability of certain carbon and nitrogen sources and also production of other metabolites. Furthermore, *P. aeruginosa*

isolates from different CF patients have been reported to show a broad range of phenotypes and growth characteristics *in vivo*, such as planktonic growth or biofilm formation [149] presumably providing basis for the selection strategy.

It has been postulated that the lower airways in chronically infected individuals represent a complex ecosystem where infections are driven by communities themselves by constant interplay between the host immune responses and a pathogen as well as between bacterial species [148]. Kelpac-Ceraj *et al.* suggested that looking at the composition of this ecosystem would allow better prediction of disease progression than investigating specific pathogens [150]. In order to identify the key selective drivers in this process the secretion characteristics of the whole community was examined using an untargetted -omics approach, as it requires determination of changes in the levels of a large number of metabolites. ^{1}H NMR spectroscopy has been used previously to characterise *Pseudomonal* growth (PAO1 type strain) in a standard laboratory medium [156]. Here, the study is to be repeated with *Pseudomonal* CF isolates from patients cultured in a synthetic airway model medium. The challenge for the method will be the fact that this study investigates differences at the species level and the composition of bacterial communities in CF individuals shows relatively low variation, therefore the nutritional characteristics of their secretions might exhibit only subtle changes that only a sensitive method can detect. It was hypothesised that NMR metabolomics of isolates from sputum samples alone from cystic fibrosis patients will allow discovery of metabolic classifiers that would enable assessment of the severity and course of the infection. The study investigated metabolic differences between different *Pseudomonal* isolates, while aiming to more closely replicate the physiochemical composition of CF airway secretions in a controlled manner and cultured clinical isolates in a defined synthetic CF medium (SCFM).

3.2 Materials and methods

Here, *P. aeruginosa* strains isolated from sputum samples collected from 13 adult CF patients were subjected to ^{1}H NMR metabolomics analysis. For patients 12 and 13 two dominant morphotypes were selected. Table 3.1 on the following page lists the details of isolates, phenotypic characteristics, and corresponding patients.

Table 3.1: ***Information on isolates, the patients that they were obtained from, and properties of the sputum sample, from which the isolate is derived.*** *Diversity-indicates the number of bacterial species identified by 16S rRNA gene clone sequence analysis; CFPE-cystic fibrosis pulmonary exacerbation; genotype I was phe508del for all patients; cfu/ml equiv. refers to mean P.aeruginosa cells numbers per ml of spent medium, as determined through Q-PCR enumeration. Data from Dr Geraint B. Rogers.*

isolate	metabolomic cluster	mucoidy	pigmentation	auxotrophy	age	sex	genotype II	BMI	diabetes	mean FEV_1 % pred.	CFPE in 12 months	diversity	culture pH	cfu/ml equiv.
CF1	I	mucoid	yes	methionine	30	male	unknown	29.0	no	54.9	3	28	5.32	$1.17x10^9$
CF2	IIc	non-mucoid	no	no	45	female	unknown	18.5	yes	40.2	4	2	6.54	$8.69x10^8$
CF3	I	non-mucoid	no	proline	47	male	unknown	20.7	yes	33.9	0	5	6.32	$2.29x10^{10}$
CF4	IIc	non-mucoid	no	no	30	female	711+3A7G	25.0	no	38.2	3	4	6.62	$4.05x10^9$
CF5	IIc	non-mucoid	no	no	22	female	phe508del	19.0	no	36.2	5	6	6.61	$2.03x10^{10}$
CF6	-	non-mucoid	yes	no	55	male	G85E	24.5	no	52.2	2	24	6.26	$1.34x10^8$
CF7	IIa	non-mucoid	yes	no	21	female	phe508del	19.0	no	56.6	4	2	6.19	$1.58x10^9$
CF9	IIb	non-mucoid	yes	no	22	male	phe508del	17.9	yes	16.5	7	8	6.93	$2.40x10^9$
CF10	I	non-mucoid	no	no	18	female	phe508del	22.5	no	84.0	3	17	5.49	$5.25x10^9$
CF11	IIa	non-mucoid	yes	no	24	female	G542X	23.4	no	72.5	3	5	6.06	$7.80x10^9$
CF12	IIb	mucoid	yes	no	20	male	phe508del	30.4	no	26.8	4	14	6.53	$5.25x10^9$
CF12b	I	mucoid	no	no	30	male	phe508del	30.4	no	72.5	4	14	5.83	$1.76x10^9$
CF13	IIb	mucoid	yes	no	20	male	phe508del	21.0	no	54.4	4	6	6.65	$1.81x10^9$
CF13b	IIa	mucoid	yes	no	21	male	phe508del	21.0	no	26.8	4	14	6.87	$8.55x10^9$
CF14	IIc	non-mucoid	yes	methionine	23	male	phe508del	20.7	yes	53.6	4	42	6.58	$2.18x10^9$

Table 3.2: ***Relationships between the sample characteristics and strain cluster membership***. *Assessment of significance was performed using a one-way ANOVA unless stated (#) whereby a Kruskal-Wallis test was used. Asterisk denotes significant ($p < 0.05$). R^2 indicates the amount of variance in the characteristics accounted for by the cluster membership. Data from Dr Geraint B. Rogers.*

Characteristic	p-value	R^2
Mucoid#	0.912	0.041
Pigmentated#	0.277	0.295
Auxotrophy#	0.757	0.089
Age#	0.361	0.212
Sex#	0.188	0.368
BMI	0.429	0.232
Diabetes#	0.572	0.154
Mean FEV_1	0.016*	0.629
CFPE	0.162	0.388
Species richness#	0.284	0.159
Culture pH	0.004*	0.751
Sputum pH	0.078	0.479
cfu/ml#	0.697	0.124
Firmicutres	0.337	0.276
Fusobacteria#	0.599	0.185
Actinobacteria#	0.747	0.114
Proteobacteria	0.350	0.269
Bacteriodetes	0.441	0.227

Table 3.3: ***Summary of the significant ($p < 0.05$) pairwise Spearman's correlation coefficients observed between the sample characteristics.*** *All other correlations were found to be non-significant. Data from Dr Geraint B. Rogers.*

Characteristic	Rho
Mucoid: Sex	-0.645
Mucoid: BMI	0.539
Mucoid: Species richness	0.652
Mucoid: Firmicutes	0.782
Mucoid: Proteobacteria	-0.556
Sex: Species richness	-0.559
BMI: Diabetes	-0.630
BMI: Firmicutes	0.691
Mean FEV_1: Culture pH	-0.736
CFPE: Culture pH	0.573
Species richness: Firmicutes	0.843
Species richness: Fusobacteria	0.660
Species richness: Proteobacteria	-0.955
Species richness: Bacteriodetes	0.693
Firmicutes: Proteobacteria	-0.743
Fusobacteria: Actinobacteria	0.900
Fusobacteria: Proteobacteria	-0.750
Fusobacteria: Bacteriodetes	0.761
Actinobacteria: Proteobacteria	-0.573
Actinobacteria: Bacteriodetes	0.585
Proteobacteria: Bacteriodetes	-0.823

3.2.1 Bacterial growth conditions

Bacterial quantification and genotyping in samples has been performed by Dr Geraint B. Rogers. *P. aeruginosa* density in samples at harvesting was determined by quantitative (Q) PCR enumeration of *opr*L gene copies in total DNA extracts, using a protocol described previously [157]. Random Amplified Polymorphic DNA (RAPD) assays were performed for each *P. aeruginosa* isolate as described previously [158]. A defined synthetic CF medium (SCFM) closely replicating the physiochemical composition of CF airway secretions was used as described previously [159–162]. The SCFM contained: 10 g/L BSA, 10 g/L porcine gastric mucin, 1.4 g/L herring sperm DNA, 10 mM MOPS, 5 g/L egg yolk emulsion, 3.6 μM $FeSO_4$, 51.8 mM NaCl, 2.28 mM NH_4Cl, 2.128 mM L-lysine HCl, 14.9 mM KCl, 1.78 mM L-alanine, 1.754 mM $CaCl_2$, 1.661 mM L-proline, 1.609 mM L-leucine, 1.549 mM L-glutamate HCl, 1.446 mM L-serine, 1.3 mM NaH_2PO_4, 1.25 mM Na_2HPO_4, 1.203 mM L-glycine, 1.12 mM L-isoleucine, 1.117 mM L-valine, 1.072 mM L-threonine, 0.827 mM L-aspartate, 0.802 mM L-tyrosine, 0.676 mM L-ornithine HCl, 0.633 mM L-methionine, 0.606 mM $MgCl_2$, 0.53 mM L-phenylalanine, 0.519 mM L-histidine HCl, 0.348 mM KNO_3, 0.306 mM L-arginine HCl, 0.16 mM L-cysteine HCl, 0.119 mM diethylene triamine pentaacetic acid, 0.013 mM L-tryptophan. The pH of the medium was adjusted to 6.8. The medium was filter sterilised using a 0.45 μm-pore-size syringe filter with the exception of porcin gastric mucin, which was sterilised separately by heating at 70 °C for 24h in 95 % ethyl alcohol as described previously [163]. Figure 3.3 on page 63 shows representative ^{1}H NMR spectra of selected constituents.

Incubation was performed in 9 ml volumes of SCFM in 15 ml Falcon tubes (BD Biosciences, Oxford, UK) with tight lids, for 72 hours at 37 °C , with inversion every 12 hours. Following incubation, bacterial cells were pelleted by centrifugation at 12,000 x *g*, 10 min at 4 °C , with the supernatant transferred to fresh NMR tubes with 10 % v/v D_2O added to provide a deuterium lock signal.

3.2.2 NMR

^{1}H NMR spectra were recorded on a Bruker Avance 400 MHz spectrometer equipped with a 5 mm QNP probe (Bruker UK Limited, Coventry, UK) with sample isolates tested in triplicate and kept at room temperature. A zgesgp pulse sequence (Bruker) with excitation sculpting using gradients was used [164]. The ^{1}H 90 degree pulse was 9.75 μs. For each spectrum, 65,536 data points were

acquired with 16 scans. To help in the assignment of the metabolite resonances, J-resolved 2D correlation with presaturation during relaxation delay using gradients (JRES, Bruker) spectra were recorded for some of the samples, using default pulse sequences as provided by Bruker. The spectral width was 20 ppm. Free induction decays were multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. The spectra were Fourier transformed and calibrated with 2,2,3,3,-D4-3-(Trimethylsilyl) propionic acid sodium salt (TSP-2,2,3,3-D4) with reference signal at 0 ppm. Phase correction was performed manually and automatic baseline correction was applied.

3.2.3 Multivariate analysis

A detailed explanation of the pre-processing and processing methods and parameters is discussed in Chapter 2 on page 18 and the software manual can be found in Appendix A on page 177.

Regions above 9.074 ppm and below 0.116 ppm were excluded because of excessive noise content and few signals. The water peak, ethanol and TMSP reference signal were also excluded. The spectra were bucketed using a 0.02 ppm bin size with additional, manual bucketing applied to adjust for peak shifting as described below, leaving 336 data points per spectrum. Spectra were normalised using probabilistic quotient normalisation (PQN) [124]. PCA was used to identify clustering patterns from the major variations between the 49 NMR spectra. For this analysis, spectra were Pareto scaled after normalisation. In order to find metabolomic classifiers for each cluster membership and provide a robust statistical analysis of the models, each possible PCA cluster was analysed against SCFM cluster using orthogonal PLSDA (OPLSDA) in a series of binary comparisons subjected to cross-validation procedure. Here, spectra were autoscaled. Both normalisation and autoscaling were included in the cross-validation. In the process 75 % of the samples were used as a training set and the remaining 25 % as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1-score used to choose the number of components, with the additional constraint to use a maximum of 8 components. This was repeated 2,000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This leads to 4×2000 models. The same procedure was repeated with randomised predictor

variables (**Y** table) to provide a reference Q^2 value. Resonances identified as significant from backscaled loadings were verified against the peak intensity of the original spectra after PQN normalisation. Peaks providing the basis for discrimination between the classes were assigned by comparing chemical shift values and multiplicities from J-resolved NMR spectra to values from the Biological Magnetic Resonance Data Bank (BMRB) [92] and The Human Metabolome Database (HMDB) [91], by analysis of published *P. aeruginosa* metabolic data [165, 166] and NMR spectra generated from individual medium components (Figure 3.3 on page 63).

3.2.4 Relationships between PCA and clinical characteristics

This statistical analysis was performed by Dr Damian Rivett using R (v.2.13.0, www.r-project.org). One-way factorial ANOVA were performed to test for significant relationships between the *P. aeruginosa* strain cluster membership and sample characteristics, with a significance threshold of $p < 0.05$. Homogeneity of variance and normality of errors were assessed using the Fligner-Killeen and the ShapiroWilk tests respectively prior to the ANOVA. If a factor failed either test a nonparametric Kruskal-Wallis rank sum test was performed. Factors that were found to be significant using ANOVA were further studied using Tukey's honest significant difference (HSD) as a *post hoc* test. Correlations between the sample characteristics were performed using Spearman's rho correlations.

3.3 Results

3.3.1 ^{1}H NMR spectroscopy of *Pseudomonas* CF isolates cultured in an airway model medium

^{1}H NMR spectroscopy has been used previously to investigate the growth of *P. aeruginosa* type strain PAO1 in Luria-Bertani broth, a standard laboratory medium [156]. In that study NMR was used to demonstrate metabolic differences between planktonic and biofilm modes of growth as reflected by the composition of the spent medium. This study was a step further and investigated metabolic differences between different *Pseudomonal* isolates, while aiming to more closely replicate the physiochemical composition of CF airway secretions in a controlled manner and cultured clinical isolates in SCFM.

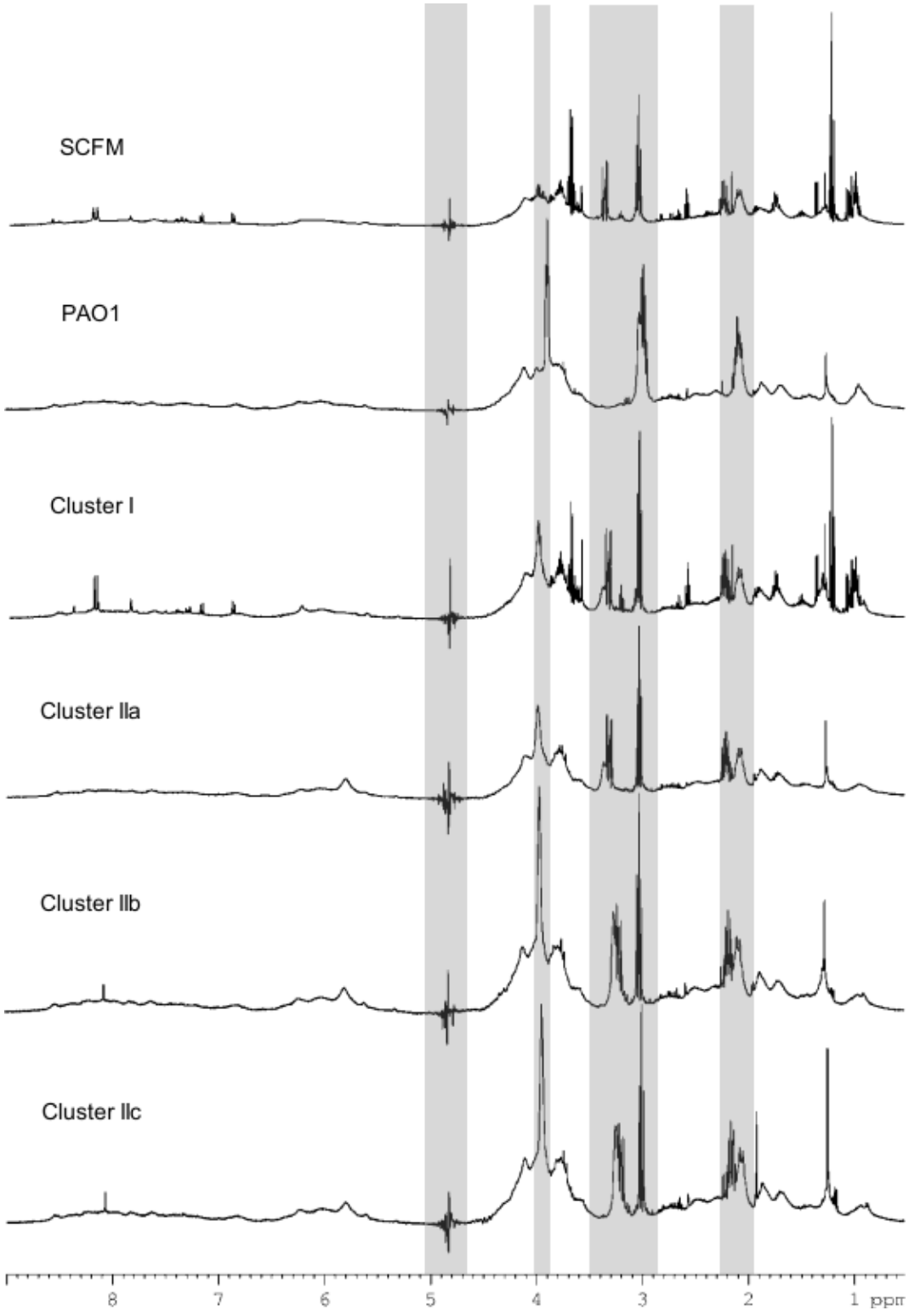


Figure 3.2: ***Representative ^{1}H NMR spectra generated from non inoculated SCFM, PAO1 inoculated SCFM, and representative members of each of the four putative clinical isolate clusters.*** *Shaded regions indicate large regions of the ^{1}H NMR spectra excluded on the basis of solvent or buffer peaks-water around 4.8 ppm and MOPS peaks around 2.2, 3.0-3.3 and 3.9 ppm.*

Representative 1D ^{1}H NMR spectra are shown for each of the isolate clusters identified by PCA described below, revealing the effect of culturing either *P. aeruginosa* PAO1 or CF clinical isolates (Figure 3.2 on the previous page). Isolate 6 was excluded from analysis due to insufficient growth in SCFM medium.

Methodological problems

^{1}H NMR spectra generated for each isolate and for the sterile medium showed high degree of reproducibility within replicates. This study successfully identified metabolic changes and related them to univariate measures of patient outcomes, however adherence to the SCFM growth protocol was problematic for the NMR study and subsequent multivariate analysis. Since liquid-state NMR was used, large complexes, viscous or solid constituents in the sample will give rise to broad peaks, which obscure the signal from metabolites in the NMR spectrum. This problem is discussed in Chapter 2 in Section 2.2.2 on page 24 and one of the solutions for the future study is to use cpmg pulse sequence [78]. SCFM contains a number of components, such as mucin and BSA, which resulted in peak broadening and loss of resolution due to high viscosity (Figure 3.3 on the next page). In addition, a number of the strains analysed were highly mucoid and capable of producing large amounts of exopolysaccharide (EPS), which leads to the same problem. Finally, although less noticeable in the SCFM spectrum, resonances attributable to the, presumably, non-metabolised MOPS buffer dominate the spectra derived from spent media (Figure 3.4 on page 64).

These components of the sample led to a number of very broad resonances, particularly between 3.5 and 4.5 ppm, and substantial, pH dependent shifting of both broad and sharper resonances between 3.00 and 3.30 ppm and around 2.10 ppm, (Figure 3.2 on the previous page) as expected from the pH dependence of buffer resonance chemical shifts (Figure 3.5 on page 65). Noticeable peak shifts were most likely due to pH- and ionic strength-induced alteration to the ionisation equilibrium for the function groups of the MOPS buffer. To remove the influence of MOPS buffer from the analysis, the following regions were excluded from further analysis in addition to the water and ethanol peaks (5.024.65 ppm; 3.703.76 and 1.221.15 ppm): 4.023.87 ppm; 3.502.87 and 2.271.96 ppm. Large scale shifting of buffer resonances resulted in ineffective peak alignment using correlation optimised warping (COW) (Figure 3.6 on page 66). Although peak realignment could conceivably be achieved through pH adjustment of spent

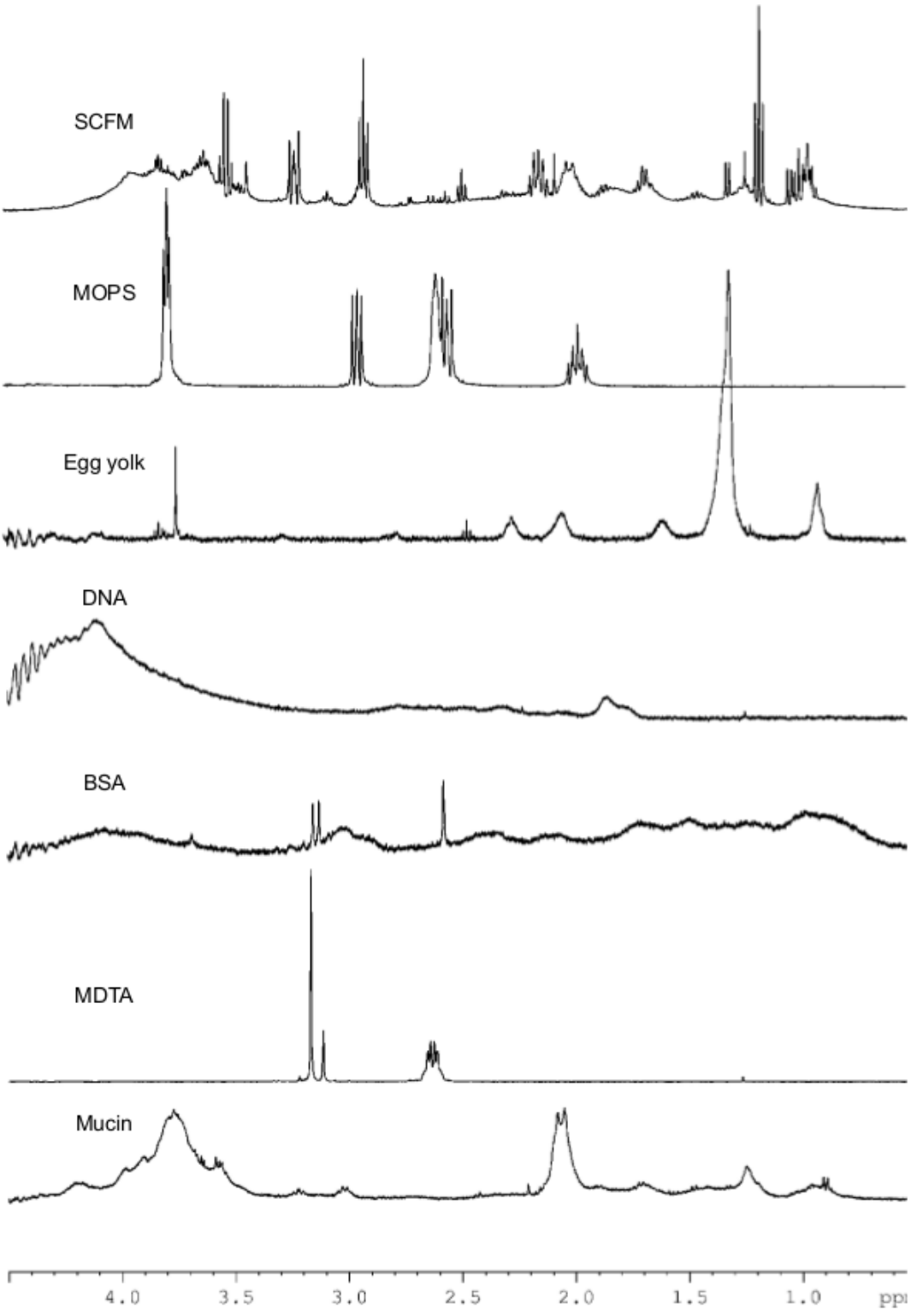


Figure 3.3: ^{1}***H NMR spectra generated for SCFM media and selected constituents.***

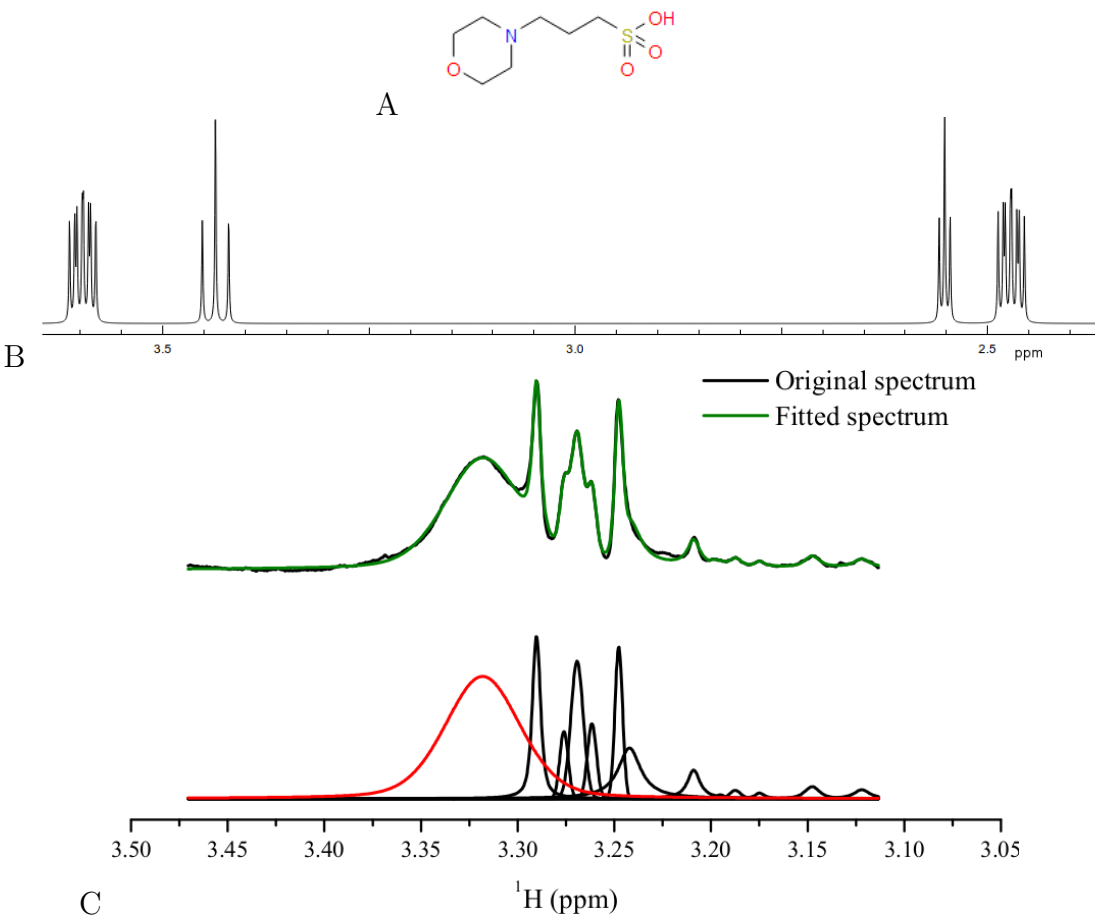


Figure 3.4: ***Chemical structure of MOPS buffer (A), simulated 1D ^{1}H NMR spectrum of MOPS buffer (B) and fragment of NMR spectrum showing the problem of obscuring broad MOPS buffer peak (C).***

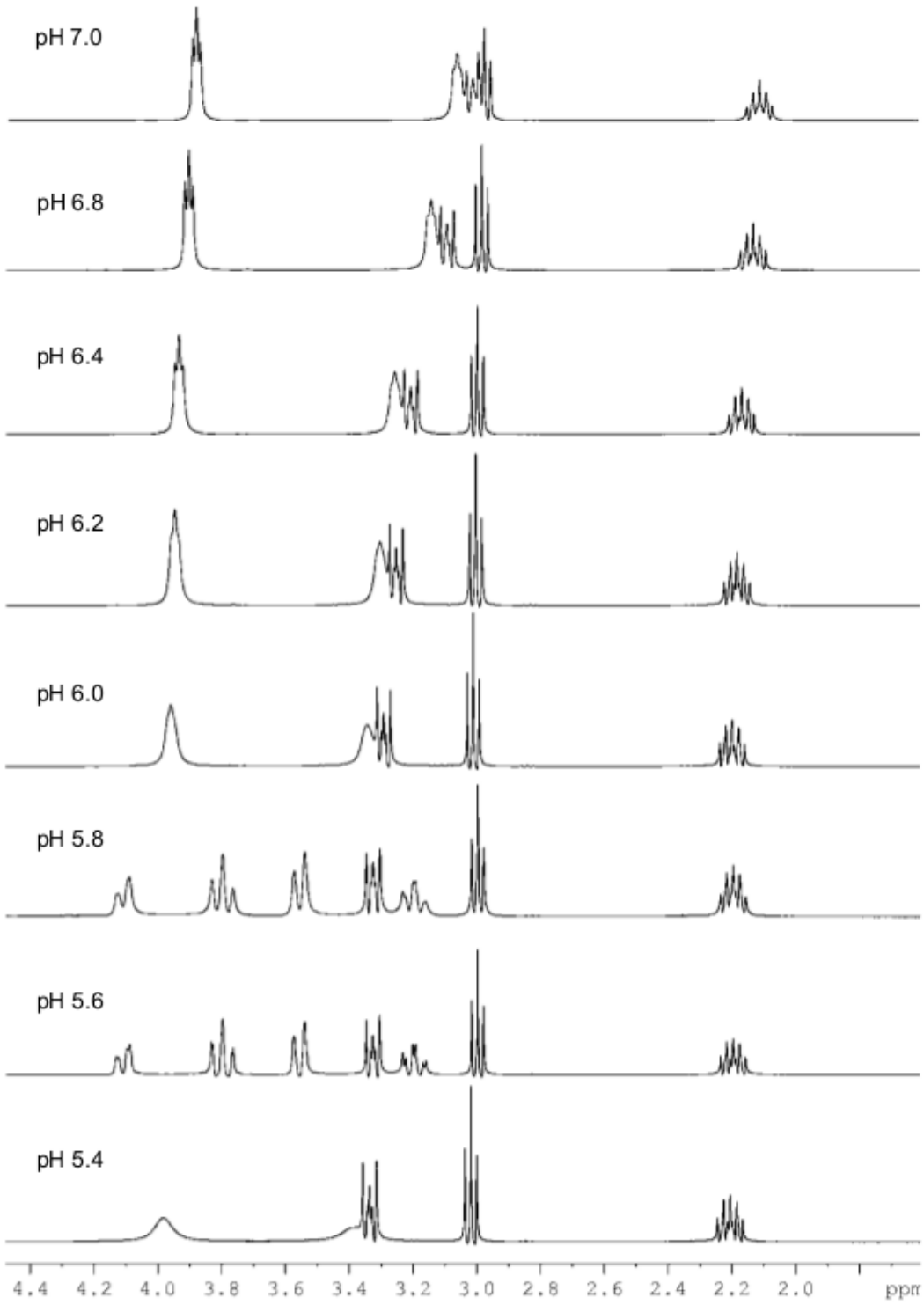


Figure 3.5: ^{1}***H NMR spectra generated for MOPS buffer in 10% D_2O at various pH.***

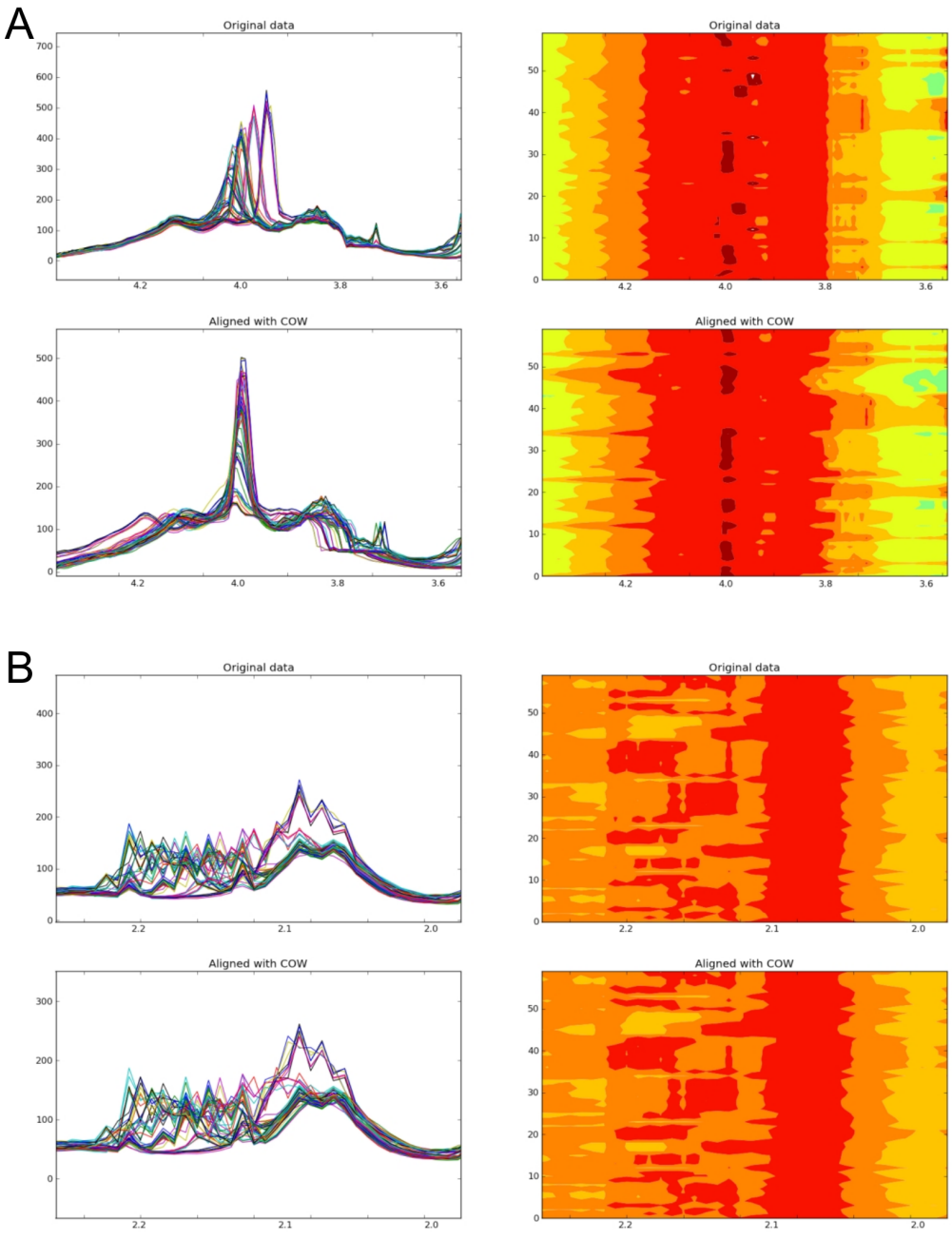


Figure 3.6: ***Correction of peak shifting using Correlation Optimised Warping (COW).*** *Correction works well for the region containing between 4.20 and 3.60 ppm-MOPS buffer (A), however alignment of the multiple peaks between 2.30 and 2.00 ppm is unsuccessful-MOPS buffer (B).*

media as is done for e.g. urine samples [167], manual peak bucketing was able to eliminate the observed peak shifts. The exclusion of large regions, considerable spectral overlap and the appearance of broad resonances following bacterial growth also precluded the use of statistical correlation spectroscopy (STOCSY) or other two-dimensional techniques to aid assignment for many resonances. For this reason our present efforts were largely restricted to using the ^{1}H NMR technique and multivariate analysis to divide the isolates into groups with similar growth strategies.

3.3.2 PCA identifies putative clusters

The clustering patterns between spectra obtained for the clinical isolates (n = 41) and SCFM medium (n = 8) were identified using PCA. The representative 2D scores plots of component 1 (PC1) versus component 2 (PC2), which explain 67.7 % of the variation in the spectra, reveal four putative separate clusters representing the different biochemical composition of the samples as detected by the NMR spectra (Figure 3.7 on the following page). In the PCA scores plot, each data point corresponds to one 1D ^{1}H NMR spectrum, and the reproducibility of the method was confirmed by the close arrangement of data points corresponding to replicates from each isolate (Figure 3.7 on the next page). NMR spectra of the CF clinical isolates fell into two, readily identifiable, main clusters (I and II) that were separated by PC1, which accounted for 56.3 % of the variation in the spectra. Cluster II is possibly subdivided into two or three further putative clusters (IIa-c) separated by PC2, with groups of isolates in separate quadrants of the PCA scores plot. The spectra from isolates in the three or four clusters were each well separated from those of the sterile synthetic media with the exception of isolate 1 which caused almost no change in the ^{1}H NMR spectra of the spent media. Cluster I was mostly separated from SCFM by PC2. Three of the four isolates found in Cluster I (1, 10, 12b) are notable in that they lead to a very acidic pH in the spent media (Table 3.1 on page 55). This might cause the cluster members to be distinguished purely on the basis of pH dropping below pH 6.0 which causes MOPS resonances to shift even beyond the ranges excluded above (Figure 3.5 on page 65) and could possibly influence the clustering of the isolates. A broad resonance at 4.04 ppm does appear for these isolates which can be assigned to the MOPS resonance expected in this region, however additional MOPS resonances expected between 3.50 and 3.87 ppm could not be discerned above the contribution from metabolite resonances. Cluster II was further separated from SCFM by PC1, but also along PC2 where isolates were further separated

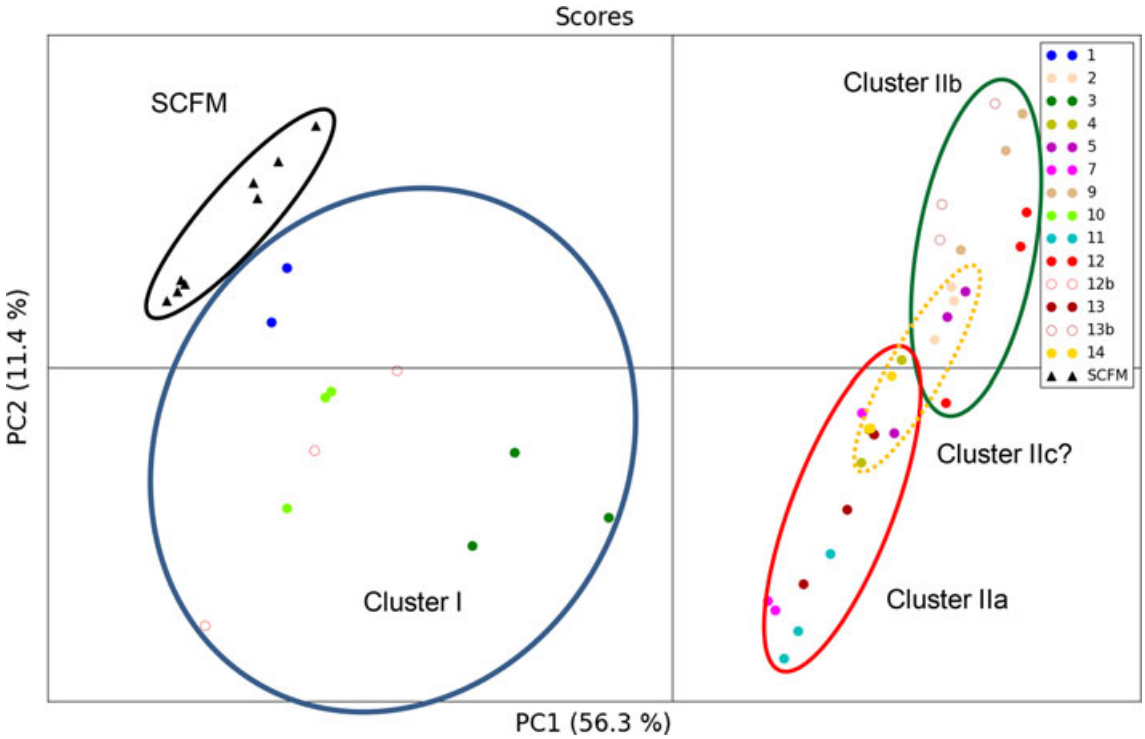


Figure 3.7: ***Scores scatter plot resulting from applying PCA to the*** ^{1}***H NMR data*** *by component 1 (PC1) and component 2 (PC2). The percentage of variance in the data explained by each component is indicated on the relevant axis. Strain identification numbers are shown. Ellipses are drawn to show putative clusters of spectra. SCFM-synthetic cystic fibrosis media.*

into upper and lower quadrants. Initially a three cluster model was considered with Cluster IIa in the lower quadrant and Cluster IIb in the upper quadrant. The existence of a fourth putative Cluster IIc, with isolates located intermediate to Clusters IIa and IIb, was considered and tested by OPLS-DA below. Key resonances whose variation contributes to PC1 and PC2 are shown in the corresponding PC backscaled loadings plots (Figure 3.8 on page 70 - Figure 3.12 on page 74) and Table 3.4 on the following page presents Q^2 values of the models resulting from each binary comparison.

3.3.3 OPLS-DA supports clusters identification

Orthogonal projection to latent structures discriminant analysis (OPLS-DA) was then used to compare ^{1}H NMR spectral data generated from sterile synthetic media with each of the four putative strain clusters and to test whether Cluster II could indeed be considered as three separate clusters. Cross-validation was performed on all models and the output plots are presented: Figure 3.8 on page 70 - Figure 3.12 on page 74. The resulting 2D scores plots show good separation

Table 3.4: ***Predictive Q^2 values for all models.*** *Cluster IIa and IIb contain isolates 4, 7, 11, 13, 14 and 2, 5, 9, 12, 13b, respectively, in the 3 cluster model and lose isolates 2, 4, 5, 14 to Cluster IIc in the four cluster model. Q^2 values for models run with permutated class assignments are given in parentheses.*

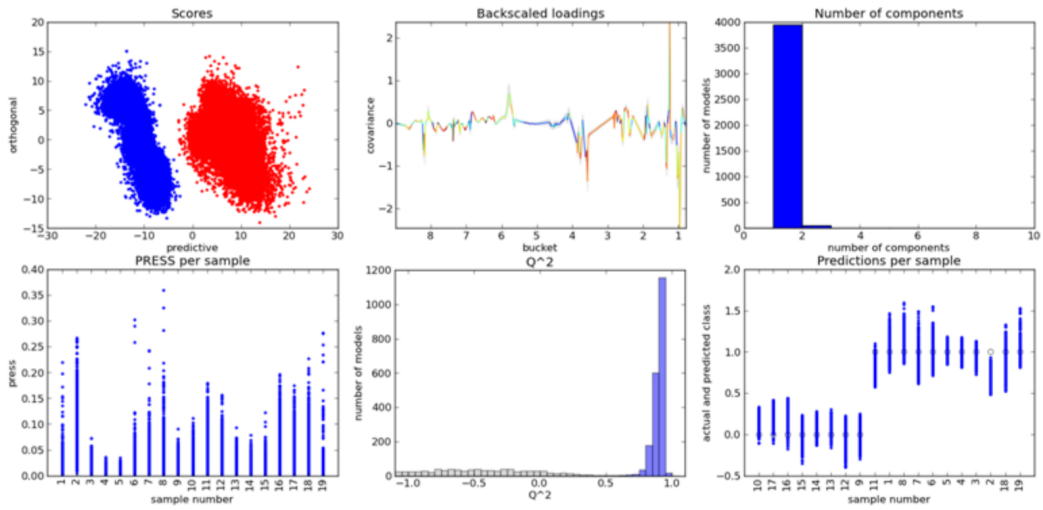
Test	3 cluster model	4 cluster model
SCFM versus cluster I	0.92 (-0.63)	0.92 (-0.63)
SCFM versus cluster IIa	0.99 (-0.44)	0.99 (-0.47)
SCFM versus cluster IIb	0.99 (-0.46)	0.99 (-0.48)
SCFM versus cluster IIc	-	0.99 (-0.40)
Cluster IIa versus cluster IIb	0.71 (-0.46)	0.91 (-0.50)
Cluster IIa versus cluster IIc	-	0.84 (-0.47)
Cluster IIb versus cluster IIc	-	0.92 (-0.44)

between the three or four putative clusters and SCFM with Q^2 values > 0.90, indicating a highly reliable model as compared with an ideal score of 1 (Table 3.4). Initially, the PCA scores plot identified two separate isolate scores clusters (Cluster I and Cluster II) with Cluster I, comprising isolates 1, 3, 10 and 12b. Using OPLS-DA, putative Clusters IIa and IIb, containing isolates 4, 7, 11, 13, 14 and 2, 5, 9, 12, 13b respectively, could be separated (Figure 3.11 on page 73) with Q^2 of 0.69. However, when a further putative cluster (IIc) was considered, comprising isolates 2, 4, 5 and 14 (PCA scores for these isolates show an intermediate distribution between the upper and lower quadrants due to PC2), the apparent separation as monitored by scores plots (Figure 3.9 on page 71 and Figure 3.10 on page 72) and Q^2 (Table 3.4) indicated that a four cluster model may be useful when greater numbers of patient isolates are available.

3.3.4 Relationships between strain cluster membership and sample characteristics

To determine whether differences in the nutritional modifications to airway secretion composition that result from the growth of *P. aeruginosa* may have clinical impacts, membership of CF sputum isolate clusters, as defined based on PCA and OPLS-DA, for both three and four cluster models, was compared with a number of potentially key strain or sputum sample characteristics. These factors were isolate auxotrophy, mucoidy, pigmentation, spent culture pH, sputum

A



B

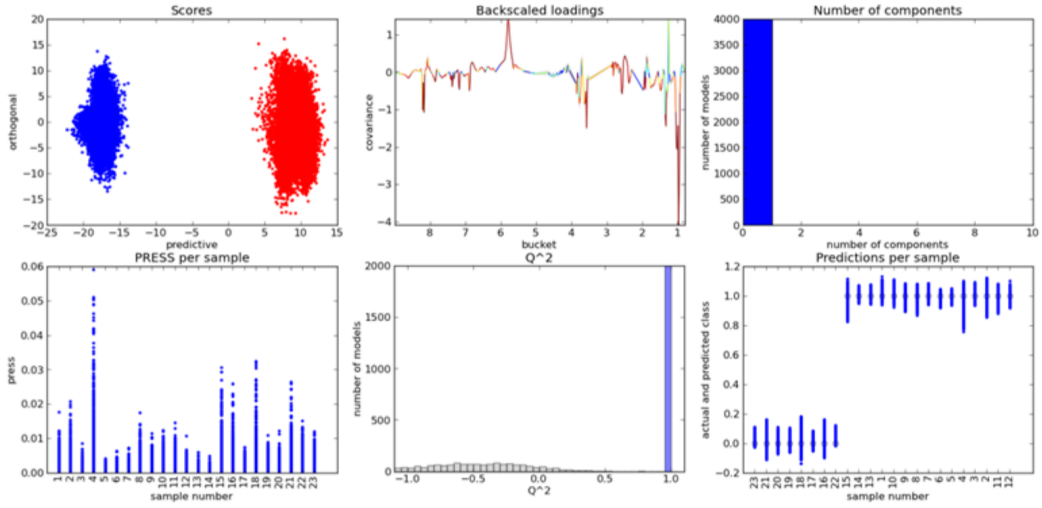
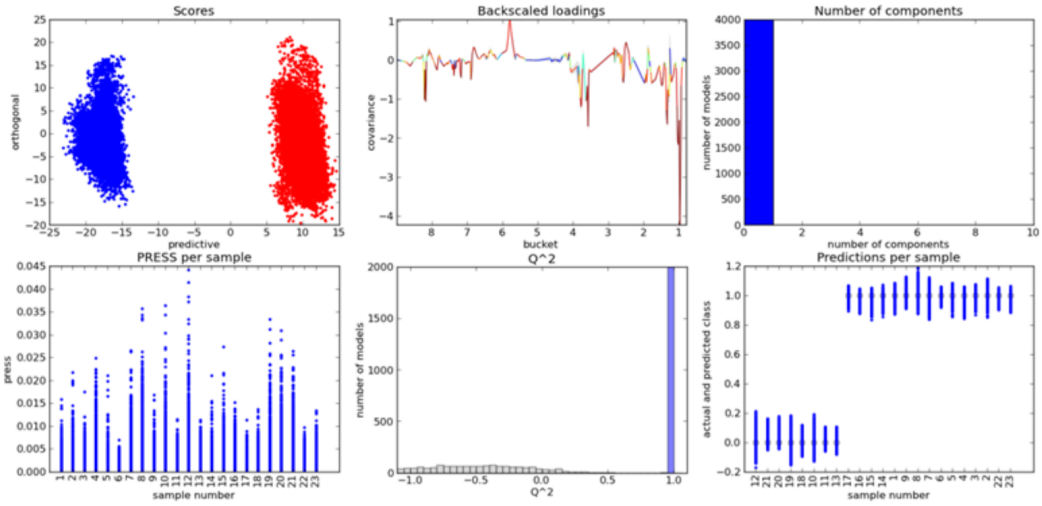


Figure 3.8: ***LOOCV output files for the comparisons of SCFM and putative Clusters I (A) and IIa (B).*** *From top left to bottom right: scores plots showing good separation between classes; back-scaled loadings plots showing* ^{1}H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram comparing random class assignment (grey) and the actual class assignment (blue); comparison of model predicted for the sample classification as compared to the actual class.*

A



B

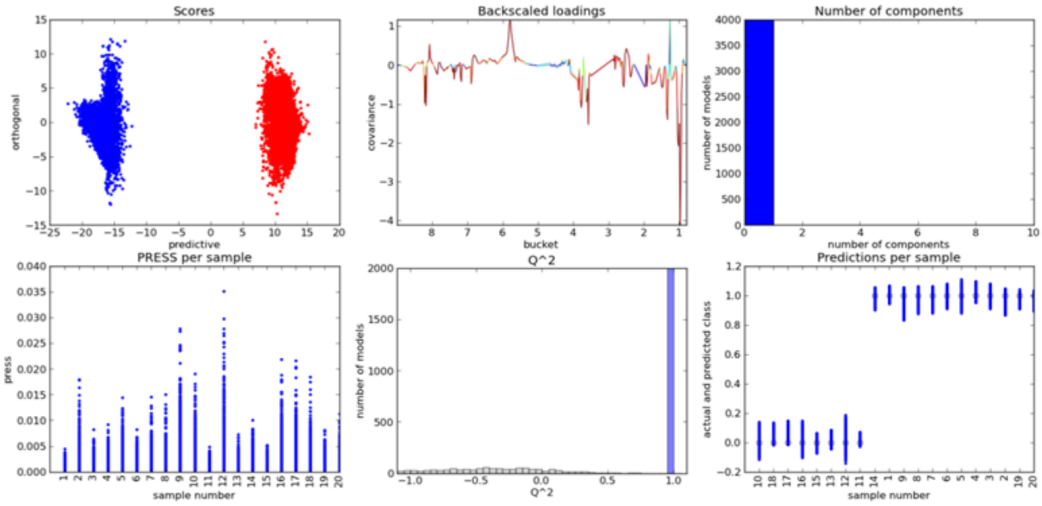
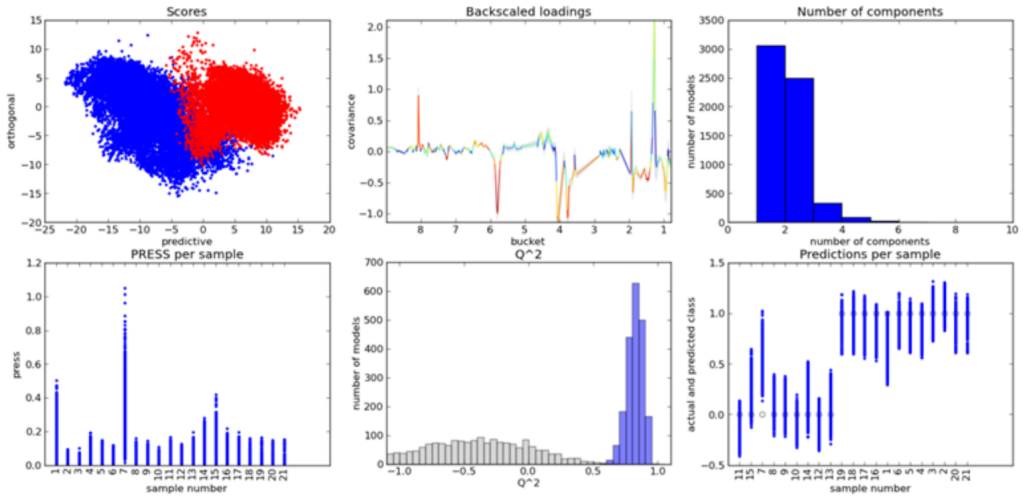


Figure 3.9: ***LOOCV output files for the comparisons of SCFM and putative Clusters IIb (A) and IIc (B).*** *From top left to bottom right: scores plots showing good separation between classes; back-scaled loadings plots showing* ^{1}H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram comparing random class assignment (grey) and the actual class assignment (blue); comparison of model predicted for the sample classification as compared to the actual class.*

A



B

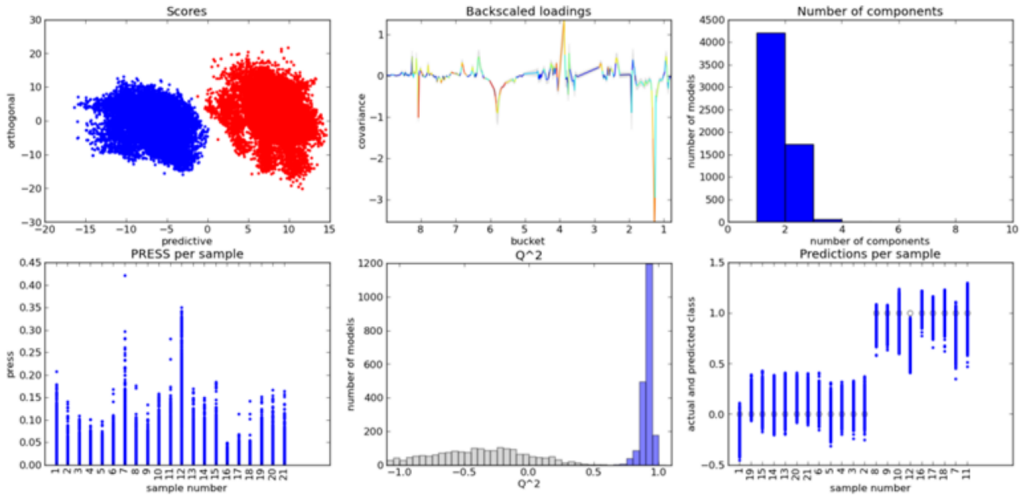
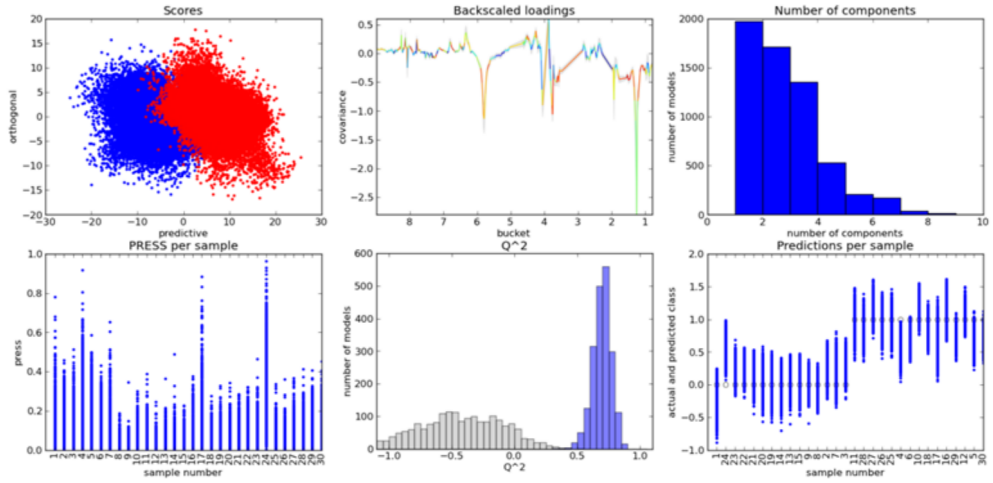


Figure 3.10: ***LOOCV output files for the comparisons of Cluster IIc and either putative Clusters IIa (A) or IIb (B) in a putative four cluster model.*** *Scores for Cluster IIc shown in red (A) and then blue (B). From top left to bottom right: scores plots showing good separation between classes; back-scaled loadings plots showing* ^{1}H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram comparing random class assignment (grey) and the actual class assignment (blue); comparison of model predicted for the sample classification as compared to the actual class.*

A



B

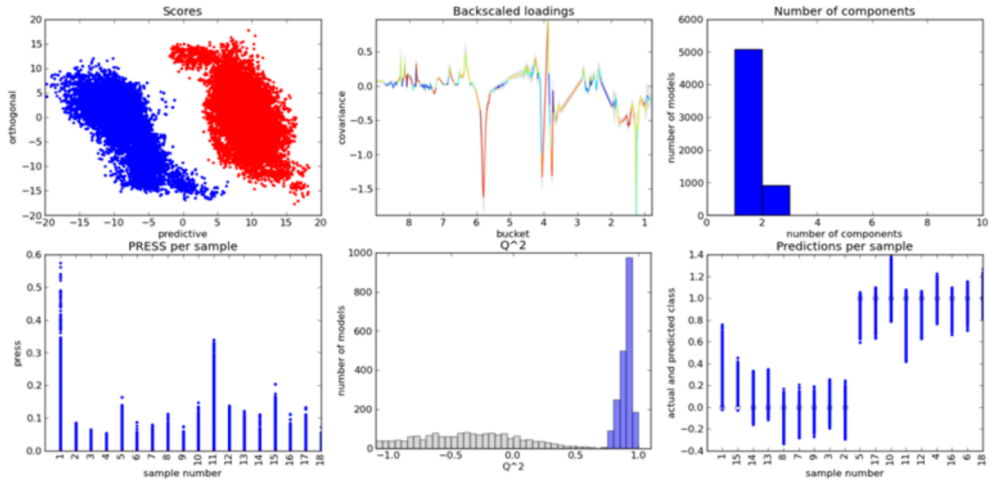
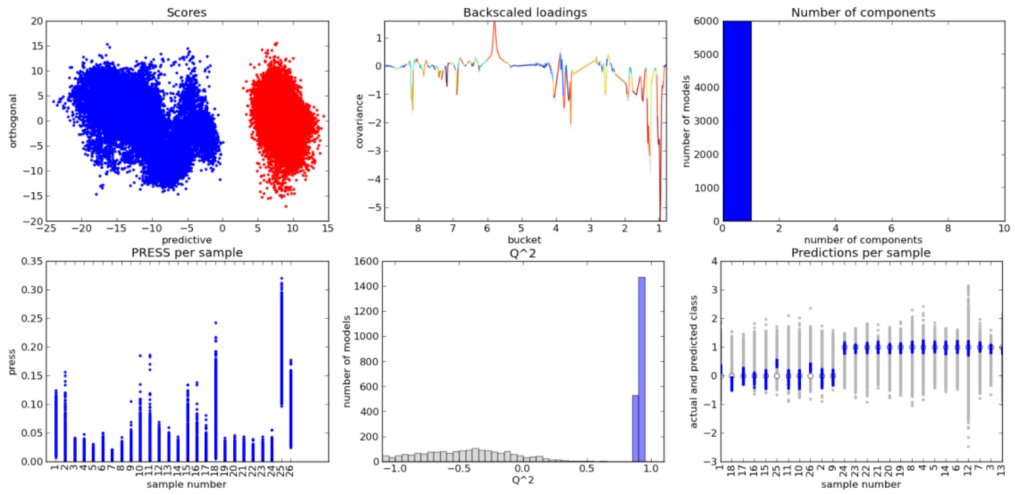


Figure 3.11: ***LOOCV output files for the comparisons of Cluster IIa (blue) and Cluster IIb (red) in a 3 (A) or 4 (B).*** *From top left to bottom right: scores plots showing good separation between classes; back-scaled loadings plots showing* ^1H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram comparing random class assignment (grey) and the actual class assignment (blue); comparison of model predicted for the sample classification as compared to the actual class.*

A



B

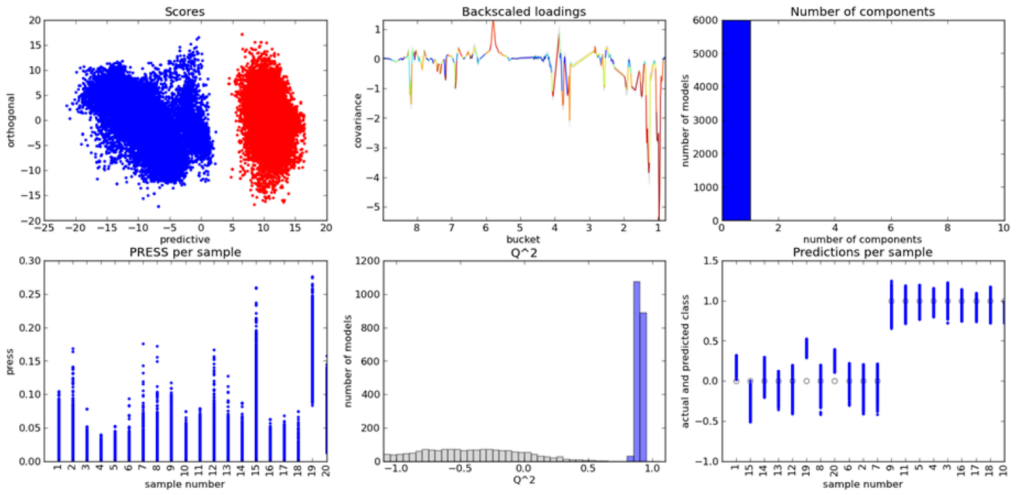
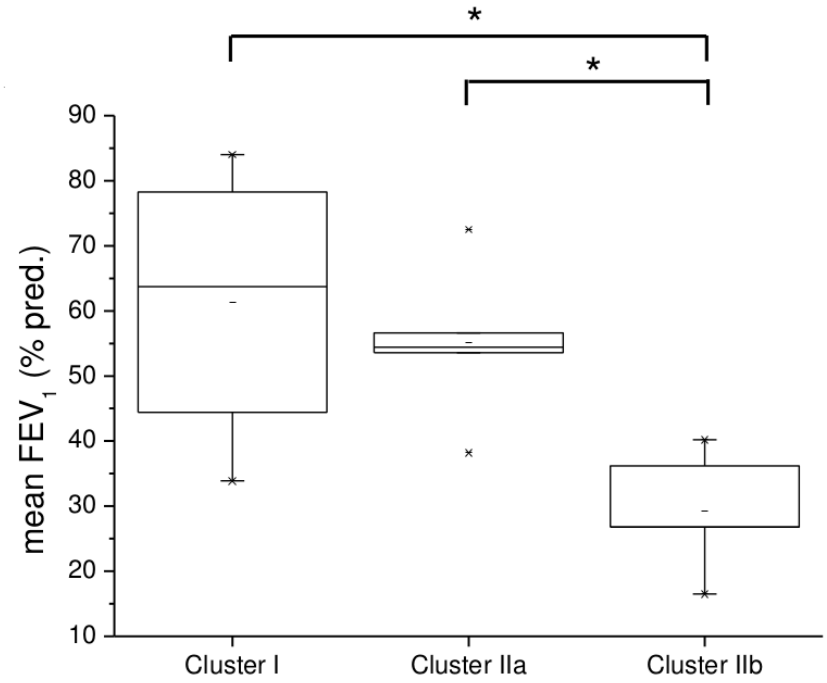


Figure 3.12: ***Leave-one-out cross-validation (LOOCV) output files for the comparisons of Cluster I (blue) and Cluster IIb (red) in a 3 (A) or 4 (B) class model.*** *From top left to bottom right: scores plots showing good separation between classes; back-scaled loadings plots showing* ^{1}H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram comparing random class assignment (grey) and the actual class assignment (blue); comparison of model predicted for the sample classification as compared to the actual class.*

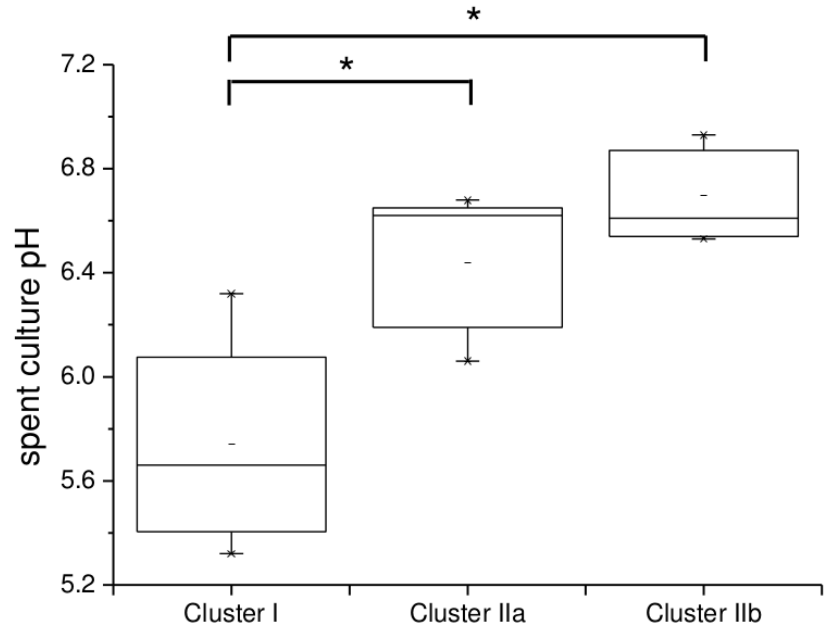
pH, species richness and relative phyla abundance as defined by 16S rRNA gene clone sequencing analysis, *P. aeruginosa* density as determined by quantitative PCR, patient age, sex, genotype, BMI, diabetic status, FEV_1, and the number of respiratory exacerbations over the preceding 12 months. This analysis has been performed by Dr Damian Rivett and the details can be found in the methods section. Highly significant relationships were identified between cluster membership and lung function (FEV_1) ($F(3,10) = 5.64$, $p = 0.0159$) and between cluster membership and spent culture pH ($F(3,10) = 8.63$, $p = 0.004$,) (Figure 3.13 on the following page). These significant relationships were tested using Tukey's honest significant difference (HSD) to assess for significant differences between clusters. In the three cluster model, this analysis found significant differences between Clusters I and IIb ($p_{adj} = 0.020$) and between Clusters IIa and IIb ($p_{adj} = 0.030$) for lung function (FEV_1) with patients in Cluster IIb having relatively poor lung function (Figure 3.13 on the next page, a). In the four cluster model, Cluster I was shown to be significantly different from both clusters IIb (padj = 0.005) and IIc ($p_{adj} = 0.010$) (Figure 3.12 on the preceding page). A possible relationship was also observed between sputum pH and cluster membership ($F(3,10) = 3.06$, $p = 0.078$, pH ranged from 5.9 to 7.8). Significant differences in spent culture pH were observed between Cluster I and both Clusters IIa and IIb in the three cluster model (Figure 3.13 on the next page, b). Therefore, although Cluster I and Cluster II are clearly separated in the PCA analysis (by PC1), the only significant differences that were found both in FEV_1 and spent culture pH were between Cluster I and Cluster IIb with a significant difference in FEV_1 also seen between Clusters IIa and IIb. These clusters are separated in the PCA analysis by PC2 and hence identification of resonances contributing to PC2, rather than PC1, or OPLS-DA analyses between these clusters should identify metabolomic changes that cause the variance in FEV_1 or spent culture pH.

3.3.5 OPLS-DA identifies characteristic metabolite consumption and production

The OPLS-DA comparisons of Cluster IIb with Cluster I (Figure 3.12 on the preceding page) and Cluster IIa (Figure 3.11 on page 73) again support the clustering determined above and allow identification of resonances from metabolites that may be implicated in FEV_1 and/or spent culture pH. Plotting normalised spectra from each of the four clusters, coloured according to cluster, highlights resonances whose intensity is consistently altered between clusters (Figure 3.15



A



B

Figure 3.13: ***Box plots comparing FEV_1 (A) and spent culture pH (B) for each of the clusters in the three cluster model; * indicates $p < 0.05$.***

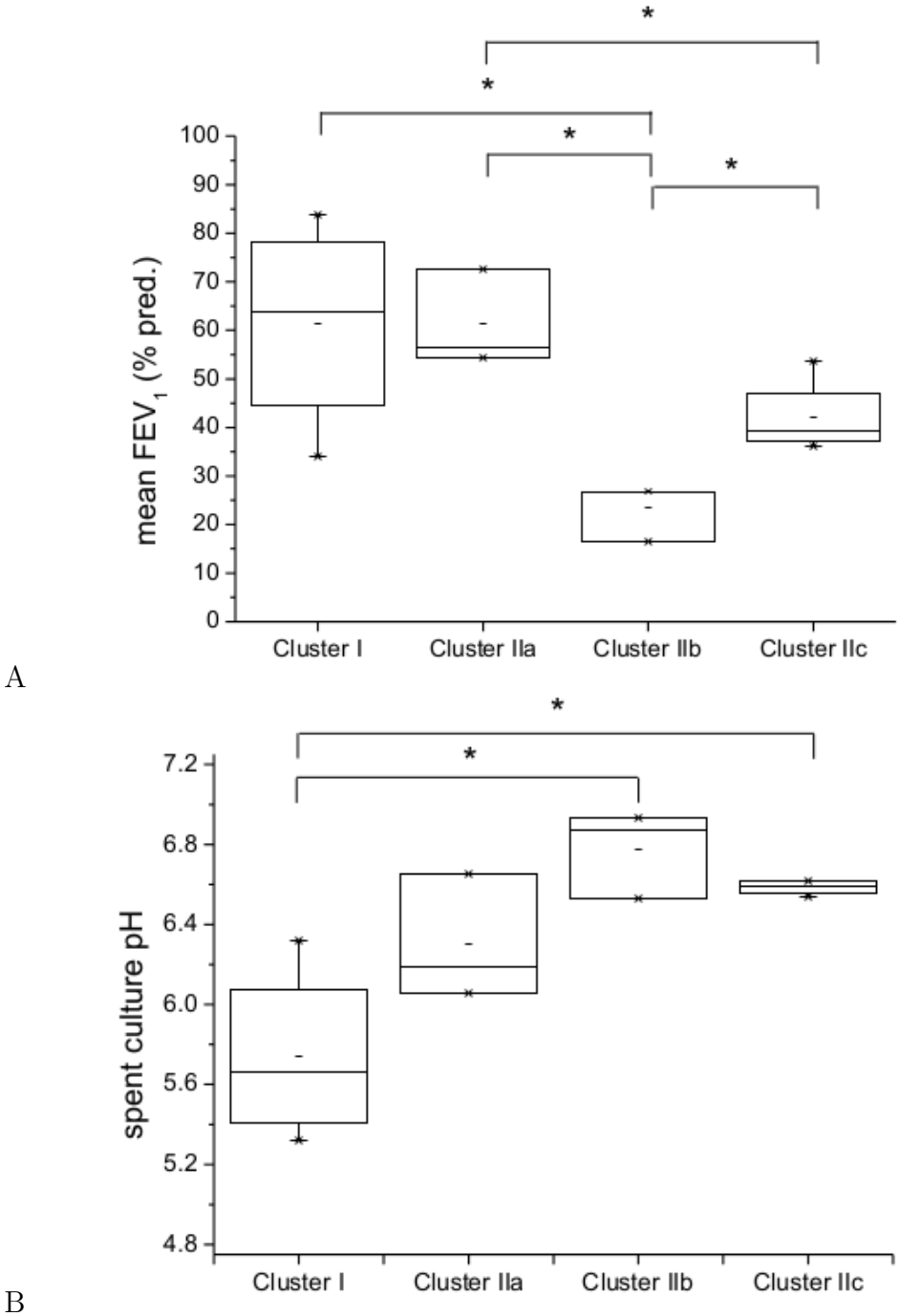


Figure 3.14: ***Box plots comparing FEV_1 (A) and spent culture pH (B) for each of the clusters in the four cluster model; * indicates $p < 0.05$.***

on the following page). Comparing the backscaled loadings plots for the comparisons between Cluster IIb and Cluster I and between Cluster IIb and Cluster IIa identifies metabolites whose differing intensities correlate with the significant differences identified above for lung function FEV_1 and/or spent culture pH. Notably, levels of lysine or ornithine appear higher in the spent media of isolates in Cluster I or IIa when compared with those from Cluster IIb, where lung function was poorest, as evidenced by their characteristic resonances at 3.765 ppm (Figure 3.15 on the next page, A) and 1.465 ppm (Figure 3.15 on the following page, B). A resonance characteristic of leucine at 1.728 ppm (Figure 3.15 on the next page, B) is also elevated in spectra of spent media of isolates in both Clusters I and IIa while a broad resonance that appears at 6.80 ppm (Figure 3.15 on the following page, C), in many spectra from Cluster IIb isolates, is largely absent from either Clusters I or IIa. Spectra from isolates in Cluster IIc are intermediate between spectra from isolates in either Cluster IIa or IIb for these features. An additional broad resonance at 5.77 ppm is notable (Figure 3.15 on the next page, C) but its intensity does not correlate with FEV_1 and is one of the main resonances that contributes to the separation of the isolates by PC1 in the PCA analysis.

3.4 Discussion

Untargeted metabolic profiling was applied to the spent media (metabolic footprinting) to compare 15 *Pseudomonal* isolates from 13 CF patients and their nutritional adaptations to chronic CF lung infections. The metabolite composition of the spent culture media was compared between isolates, patients and with the patient lung function and various sample characteristics. The influence of the nutritional sources on bacterial community composition has been discussed previously [168, 169] and high cell number of *P. aeruginosa* colonising CF lung affects the host by selectively allowing co-infection and subsequent co-colonisation by other bacterial species.

Bacteria influence the environment within the host by large- and fine-scale modifications in the utilisation of compounds such as carbon and nitrogen sources and the production of major metabolites. It was hypothesised that probing those modifications will provide an indication of the degree to which *P. aeruginosa* isolate growth might have a differential impact on CF airway secretions and to try to identify a composition favoured during e.g. exacerbations and used them for

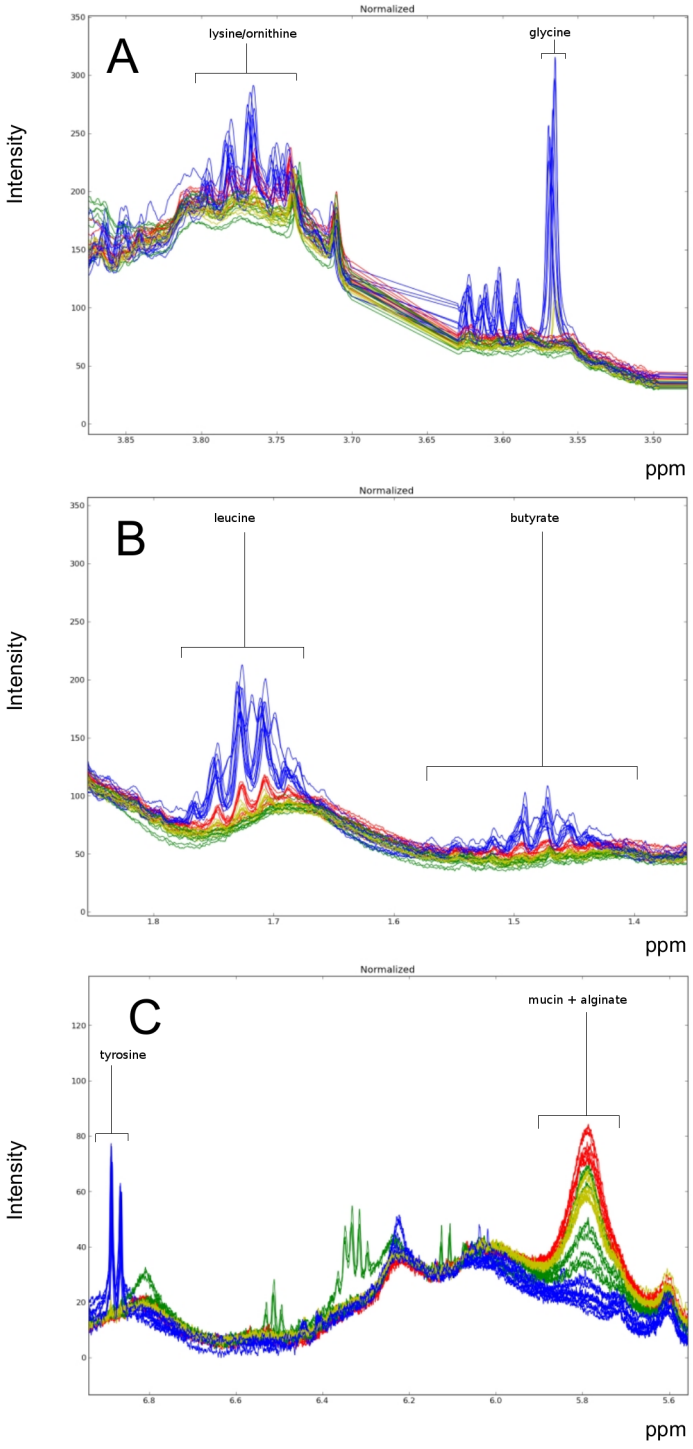


Figure 3.15: ***Normalised*** ^{1}***H spectra (with excluded regions but otherwise untreated) of spent media coloured according to cluster membership.*** *(Cluster I-blue, Cluster IIa-red, Cluster IIb-green, Cluster IIc-yellow). Spectral regions between 3.5 and 3.85 ppm (A), 5.6 and 6.9 ppm (B) and 1.4 and 1.8 ppm (C) are shown.*

early diagnosis. Insights into nutritional composition and its role in promoting or silencing communication between different species colonising CF lung and identifying the consequences of any alterations will be helpful in treatment and management of the condition.

The aim was to obtain a dynamic snapshot of major compositional changes taking a systems biology approach and the method of choice was ^{1}H NMR spectroscopy. It allows to monitor all the metabolites simultaneously with high reproducibility, therefore allowing to see the between sample variation in chemical composition.

In the experiment a defined synthetic growth medium (SCFM) was used in order to replicate more closely the composition of CF lower airway secretions. It was demonstrated in previous studies to be a suitable medium for *in vitro* model of CF lower airway conditions [160, 162]. Adherence to the well validated SCFM growth procedures presented some challenges for the NMR and in the future the constituents would have to be replaced with deuterated forms or by less viscous buffers in order to avoid pH-dependant peak shifts and peak broadening. Also, HR-MAS NMR could be considered in order to eliminate the broad resonances.

The exometabolome (spent culture media) was analysed as this provides a direct measure of the interaction of bacteria with the environment and addressed the main question: "To what degree the impact of *P. aeruginosa* growth differed between clinical isolates under conditions approximating those encountered *in vivo*". Metabolic profiling showed substantial differences between different CF isolates, which were categorised into four separate clusters (Figure 3.7 on page 68), representing four end-stage phenotypes. The clustering was driven mainly by changes in levels of amino acids, which may be an evidence of metabolic adaptation to the host environment. Strains could have evolved more efficient energy strategies in the low oxygen tensions in mucopurulent CF sputum. Studies shown that amino acid levels influence the antibiotic susceptibility of *P. aeruginosa* [170] and their growth strategies [171, 172].

Cluster membership was found to correlate with spent culture pH (Table 3.2 on page 56), despite careful exclusion of pH shifting resonances from the analysis and successful spectral alignment. However, visual inspection of the spectra from each cluster (Figure 3.2 on page 61) shows clear variation in levels of me-

dia constituents and metabolites produced/consumed between isolates and not pH-induced peak-shifts. No correlation between *P. aeruginosa* cell count and cluster membership was found (Table 3.2 on page 56), indicating that isolates do not cluster according to cell numbers, but are indeed separated by divergence in growth strategies. The cause of the change in pH of the medium could be due to change in levels of non-pH neutral components. It could also be due to proton extrusion into the medium, which is an important component of the adaptation to growth at low pH. Nevertheless, the impact of *P. aeruginosa* growth on pH could have major clinical implications since it has been shown to influence both bacterial community composition [173, 174] and behaviour [175]. Innate defence processes such as ciliary function [176] and mucus viscosity [177] have also been shown to be affected by alterations in airway secretion pH.

A highly significant relationship was also found between the cluster membership of CF isolates and the lung function of the patient from which they were obtained (as measured by FEV_1) (Table 3.2 on page 56). This demonstrates the strong relationship between the nutritional modification of the environment by *P. aeruginosa* growth and patient lung function and suggests that further research to deepen our understanding of bacterial community composition in the CF airways is needed.

The CF lung is a very heterogeneous environment with a broad range of species [147] with nutritional requirements being the main selective pressure in the colonisation process. Interestingly, bacterial species commonly found in oral cavity are not reported in CF lower airway secretions and this is thought to be due to changes in the nutritional availability which in CF lung is modified by dominant bacterial species [178, 179]. Such interactions, however, are still poorly understood in the CF lower airways.

3.5 Conclusion

^{1}H NMR was applied to see the effect of *P. aeruginosa* growth in a model CF medium on different isolates. NMR divided isolates into 4 clusters, based mainly on divergence in metabolite production, which suggests that *P. aeruginosa* isolates are able to adopt different growth strategies in response to changing environment. Highly significant relationships between *P. aeruginosa* isolate, patient lung function (FEV_1) and spent culture pH suggests future application in pre-

dicting patient lung function via characterisation of *P. aeruginosa* growth.

This study is an example of how NMR metabolomics offers a way of implementing personalised healthcare protocols. It measures the metabolite levels in an organism and models the changes in its levels in biological fluids and tissues and links those variables to health and/or diseased status. The altered levels of certain metabolites can be then used as predictors of susceptibility to disease or likely effectiveness of the treatment. Such systems (of the whole human organism) measurements and modelling approaches unravel human complexity and meet the medical requirements of today's market; the methodologies are robust, cost efficient and high-throughput. Nevertheless, this study highlights the importance of experimental planning and shows the main limitations of liquid state NMR methodology. The SCFM growth medium contains a number of large and viscous components which could not be eliminated and resulted in the presence of broad, peak-obscuring metabolic changes. The most difficult problem to overcome was pH-sensitive peaks from non-metabolised MOPS buffer. In this study peak alignment and exclusion successfully mitigated this effect, however, in future pH adjustment of spent media would be recommended as well as use of deuterated solvent forms.

4

Application of NMR metabolomics in investigation of the effects of gastrointestinal microbiota divergence in genetically identical mice

Work described in this chapter has been published as:

Rogers, G.B., Kozlowska, J., Keeble, J., Metcalfe, K., Fao, M., Dowd, S.E., Mason, A.J., McGuckin, M.A. & Bruce, K.D. Divergence in gastrointestinal microbiota in physically-separated genetically identical mice. *Scientific Reports* 2014 (4) 5437

4.1 Introduction

The microbial diversity in a human ecosystem is high and estimated to be occupied by more than 10,000 microbial species [3]. Human distal gut represents the highest density natural bacterial ecosystem known and contains more bacterial cells than all of other microbial communities within human ecosystem combined [180]. The human colon ecosystem alone has been estimated to contain more than 400 bacterial species the vast majority of which belong to two phyla of bacteria-the Bacteroidetes (48 %) and the Firmicutes (51 %) [180, 181]. These complex microbial communities are believed to contribute to health maintenance and, when in imbalance, to the development of diseases. Despite being an integral part of human biology, the development and evolution of intestinal microbiota and how its composition relates to human physiology is still considered poorly understood [160, 181, 182]. The gut microbiota has been linked to immunity [183] and nutrient intake, as well as cardiovascular health [184], obesity [29, 30, 185], diabetes [186], metabolic abnormalities [187], inflammatory bowel conditions [188], hepatic function [189] nervous system development [190], carcinogenesis [191] and recently allergy [32, 33]. Microbiome sequencing studies showed that the microbiota composition correlates with characteristic changes and health and disease risk [192, 193]. However, a new consensus emerged which suggests that bacterial community composition changes should be accompanied by studies of community function and understanding at the molecular level [194]. This can be accomplished by investigating host metabolism, modulation of host signalling pathways and synthesis of metabolites inferred from metabolomic studies. Only such a detailed understanding of host-microbiota interactions would reveal how microbiome relates to disease [194]. Metabolites are effector molecules and simultaneous analysis of a large number of metabolites in a host gut will highlight the differences in function of gut microbial communities and will allow to link them with disease characteristics which can perhaps be modulated by perturbations from small molecule drugs or pre- and probiotics [195]. Moreover, the phenotype of colonising bacteria can influence the emergent properties of the community and this in turn can be a factor modulating e.g. nutritional extraction or susceptibility to certain diseases in some hosts [26, 28]. Therefore divergence in the microbiota and metabolome of genetically-identical mice could have a significant impact on research and possibly confound the experimental findings of such studies.

Consequently researchers are becoming increasingly aware of the way gut microbial communities might influence host physiology, yet its impact on research using genetically-identical animals is not known. Using genetically homogeneous mice theoretically enables reproducible studies in animals. The genetic background of mice is stable and so is the reproducibility of that background in different laboratories and through time [196]. The objective of this study was to compare the microbiomes and the metabolomes of mice caged in four separate rooms and look for possible links between microbiota composition and various conditions.

In this chapter, ^{1}H NMR metabolomics has been applied to investigate the faecal metabolome of genetically-identical mice, housed in four separate units. Faecal samples from 20 mice housed in four separate barrier rooms, within the same facility, who were fed the same chow were used in this study. Samples consisted of individual faecal pellets taken from individual mice. After collection, pellets were placed into separate collection tubes and frozen prior to analysis.

4.1.1 Gastrointestinal microbiota composition in host

From the view point of a damage-response framework [12], a host is born sterile and the acquisition of the gut microbial flora can be considered an infection, which is later considered commensal through colonisation that does not lead to host damage due to the infection. Microbes living in the gut environment have to pass a strict selection by surviving the action of digestive enzymes, the presence of cell-surface molecular paraphernalia, attack by bacteriophages and the immune system of the host. The network of adaptive stress responses when entering a new host resulting in adaptive mutations and the ability to grow rapidly enable some microbes to survive such constant interplay of the resident gut microbial communities and the host microbiome [197, 198]. Mice provide an experimentally tractable model to investigate how resident microbial communities impact small molecule metabolites present in host biofluids and how this divergence can impact research using animal models.

4.1.2 Gastrointestinal microbiota composition discrimination in mice

16S rRNA genetic analysis, performed by Dr Geraint B. Rogers, identified bacterial species in the faecal samples obtained for this study and revealed the faecal

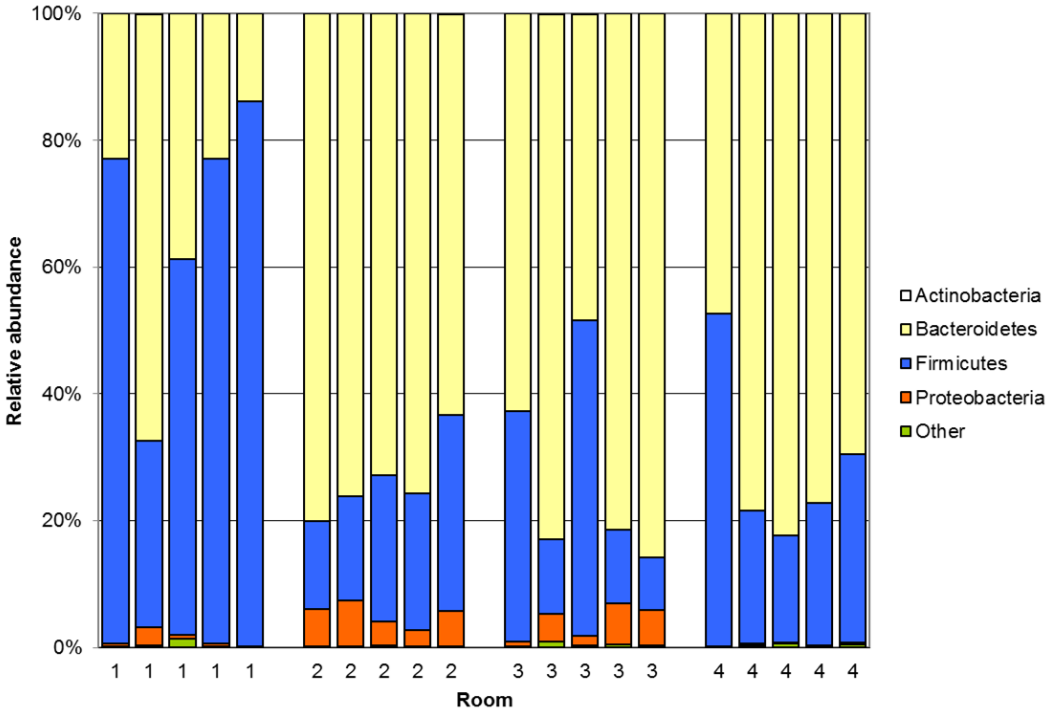


Figure 4.1: ***Variation in microbiota phyla based on the bacterial identities derived from 16S ribosomal RNA gene sequencing*** *Provided by Dr Geraint B. Rogers.*

microbiota to be dominated by the phyla Bacteroidetes and Firmicutes (62.4 ± 22.4 (SD) % and 34.7 ± 23.9 %, respectively) (Figure 4.1) with marked variation between mice. A range of diversity and richness measures were used to assess changes in microbiota composition, including taxa richness, Chao1, Shannon index, Simpson index 1-D [199] and the results are summarised in Table 4.1 on the following page. All the measures were significantly lower for mice from room 4 as compared with mice from other room groups.

Analysis of the predominant genera, shown in Table 4.2 on page 89, identified the twenty genera with the highest mean relative abundance. They were similar to genera reported elsewhere [200]. Most samples were dominated by *Prevotella*, (39.0 ± 20.2 % of sequences). Controlled ANOVA tests identified significant differences between room groups (Table B.1 on page 186). Hierarchical cluster analysis divided the samples into three clusters based on predominant genera (Figure 4.2 on page 90). Again, marked differences between room groups, but also within the groups were detected. Cluster I comprised samples from all animals from room 3 and additional animals from rooms 1 and 2, cluster II comprised all animals from room 4 and cluster III included all of the remaining animals from

Table 4.1: ***Bacterial alpha diversity*** *assessed using the Chao1 richness estimate, OTU richness, and Shannon Index, with Kruskall-Wallis test of K samples with a controlled multiple pair-wise comparison. Significant differences in measures are indicated using standard notation; samples that share a letter are not significantly different, while samples that do not share a letter are significantly different. Room 4 is significantly different than rooms 1, 2 and 3 for all measures (p <0.001). Provided by Dr Geraint B. Rogers.*

Chao1

Sample	Mean	Std. deviation	significance
Room 1	257.84	27.03	B
Room 2	282.94	58.70	B
Room 3	311.44	88.59	B
Room 4	168.43	18.52	A

OTU richness

Sample	Mean	Std. deviation	significance
Room 1	354.80	17.00	B
Room 2	421.40	151.13	B
Room 3	706.40	342.38	B
Room 4	208.00	23.67	A

Shannon index

Sample	Mean	Std. deviation	significance
Room 1	7.74	2.23	B
Room 2	10.23	1.71	B
Room 3	9.10	1.22	B
Room 4	4.45	0.38	A

Table 4.2: ***The 20 genera identified with the highest mean abundance in the faecal samples*** *collected from 20 genetically-identical mice hosted in four different rooms. Provided by Dr Geraint B. Rogers.*

Phylum	**Genus**	**Mean abundance (%)**
Bacteroidetes	*Prevotella*	38.96
Firmicutes	*Coprococcus*	16.37
Bacteroidetes	*Bacteroides*	7.07
Bacteroidetes	*Parabacteroides*	6.39
Firmicutes	*Lactobacillus*	4.38
Firmicutes	*Oscillospira*	3.82
Bacteroidetes	*Alistipes*	3.55
Bacteroidetes	*Tannerella*	3.54
Firmicutes	*Clostridium*	3.47
Firmicutes	*Roseburia*	1.91
Bacteroidetes	*Pedobacter*	1.86
Firmicutes	*Ruminococcus*	1.52
Firmicutes	*Blautia*	1.41
Proteobacteria	*Sutterella*	0.72
Bacteroidetes	*Sphingobacterium*	0.63
Proteobacteria	*Rhodospirillum*	0.50
Proteobacteria	*Novispirillum*	0.50
Proteobacteria	*Nautilia*	0.45
Firmicutes	*Butyrivibrio*	0.34
Firmicutes	*Eubacterium*	0.27

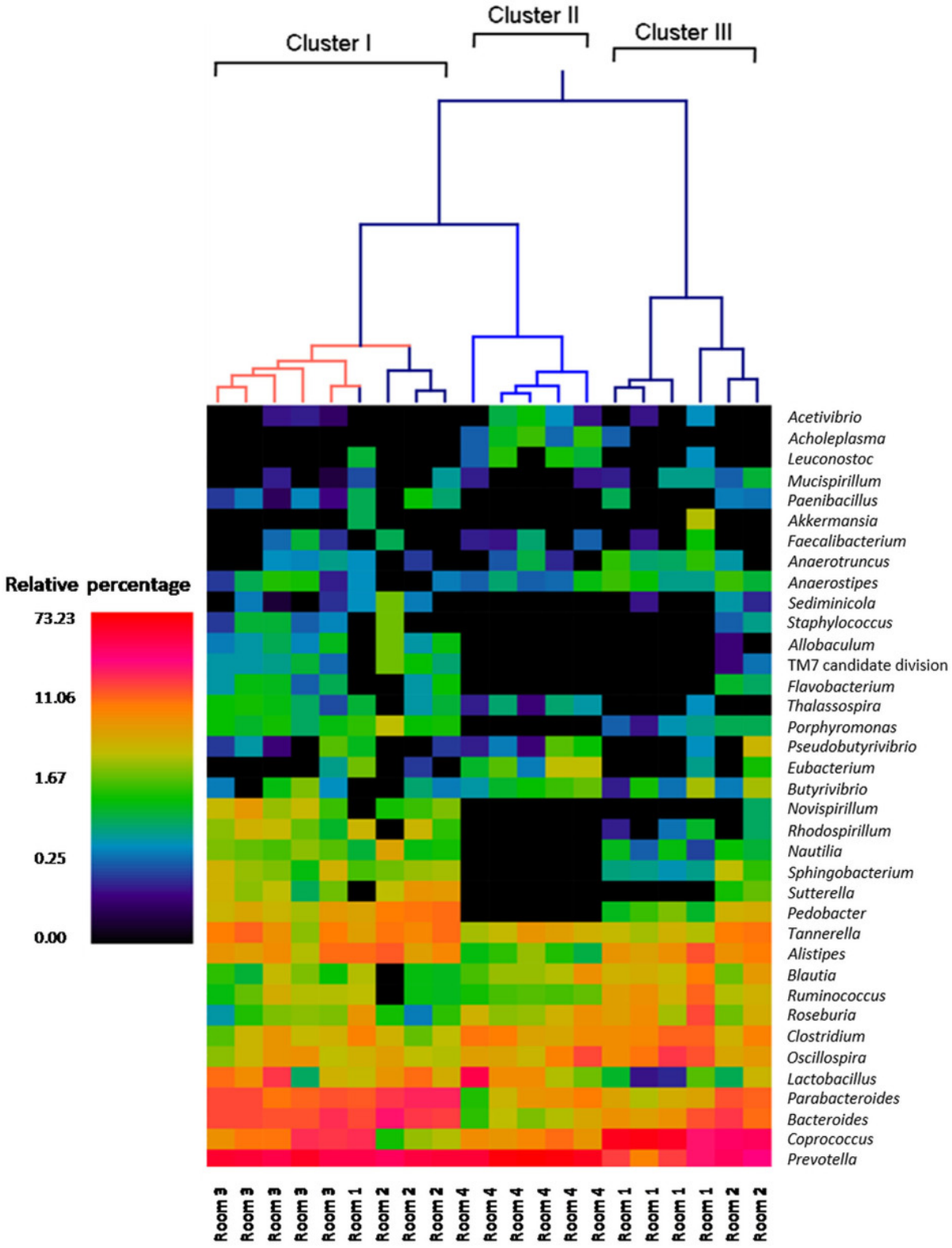


Figure 4.2: ***A hierarchical cluster diagram showing relative percentage of the associated genera in each sample****. The main differences were observed between* Prevotella, Caprococcus, Bacteroides *and Parabacteroides. Room group 1 and room group 2 exhibit some co-clustering indicating differences within the groups. Provided by Dr Geraint B. Rogers.*

rooms 1 and 2. Notable is the absence, or very low abundance, in room group 4 of a number of genera including *Sutterella*, *Sphingobacterium*, *Novispirillum* and *Porphyromonas*. Taken together, the bacterial microbiota showed marked divergence that was, in most cases, linked to room occupancy with the microbiota composition allowing classification of the samples into three clusters.

This is interesting since all mice received the same standard diet, mouse chow, which is composed of carbohydrates readily fermented in the colon to short chain fatty acids (SCFA), primarily acetate, butyrate, lactate and propionate [201, 202]. SCFA levels have been demonstrated to be important in determining the species composition of colonic microbiota, prevention of the growth of pathogens [174, 203] and modulate nutrition, adipose tissue deposition, immunity and cancer amongst other conditions [202, 204].

It was therefore investigated whether differences in microbiota composition in genetically-identical mice with identical diet could lead to distinct metabolomic characteristics and what is the overall relationship between gut microbiota and metabolome of the host. The complexity of the microbial communities, however, hampers identification of the functional connections. Studies using both mice and humans aim to show what effect the gut microbiota has on the host metabolome [198, 205–208]. Thorough understanding of the mechanisms involved in the interactions between the microbiota and its host and between metabolism of the gut microbiota and metabolic outcomes of the host are necessary to develop new treatments for metabolic diseases.

4.2 Materials and methods

4.2.1 NMR

Samples consisted of individual faecal pellets taken from individual mice. After collection, pellets were placed into separate collection tubes and frozen prior to analysis. For the NMR study, pellets of mouse faeces were resuspended by vortexing in 500 μlof phosphate buffer saline. The sample was then centrifuged at 13,000 $\times g$ for 10 min and the supernatant transferred to a fresh tube. Centrifugation was repeated and the resulting supernatant lyophilised overnight and resuspended in 500 μlof D_2O. Samples were prepared in triplicate and ^{1}H NMR spectra were acquired on 400 MHz on a Bruker Avance spectrometer (Bruker,

Coventry, UK) equipped with a 5 mm QNP probe using a zgesgp pulse sequence incorporating water suppression via excitation sculpting with gradients. The ^{1}H 90 degree pulse was 9.75 μs. The spectral width was 20 ppm. Free induction decays were multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. The spectra were Fourier transformed and calibrated to a 2,2,3,3,-D4-3-(Trimethylsilyl) propionic acid (TSP) reference signal at 0 ppm. Phase correction was performed manually and automatic baseline correction was applied. To help in the assignment of the metabolite resonances, J-resolved 2D correlation was performed with pre-saturation during the relaxation delay using gradients (J-Res, Bruker).

4.2.2 Multivariate analysis

Pre-processing and orthogonal projection to latent structures discriminant analysis (OPLS-DA) were carried out with software that was developed in our laboratory for a previous study [209] using the Python programming language with numpy and scipy for calculations, and matplotlib for visualisation. The nonlinear iterative partial least-squares (NIPALS) algorithm [108] was used for OPLS-DA analysis. Regions above 8.5 ppm and below 0.45 ppm were excluded from the analysis because of noise content. The water peak and TSP reference signal were also excluded. Spectra were bucketed using 0.005 ppm bin size leaving 1588 data points per spectrum. These spectra were normalised [124] and auto-scaled (variance of every data point normalised to 1). Cross-validation was performed where 75 % of the samples were used as a training set and the remaining 25 % as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class, and that at least one sample from each class was present in the test set. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1 score selection method was used to choose the number of components, with the additional constraint to use a maximum of 8 components. A double cross-validation was repeated 2000 times (or 100 times) with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This led to 4 $\times$ 2000 models (or 100 models). Finally, this procedure was repeated with randomly generated class assignments to provide a reference Q^2 value. Peaks provide the basis for discrimination between the classes were assigned by comparing chemical shift values and multiplicities from J-resolved NMR spectra to values from the Biological Magnetic Resonance Data Bank (BMRB) [92] and The Human Metabolome Database (HMDB) [91].

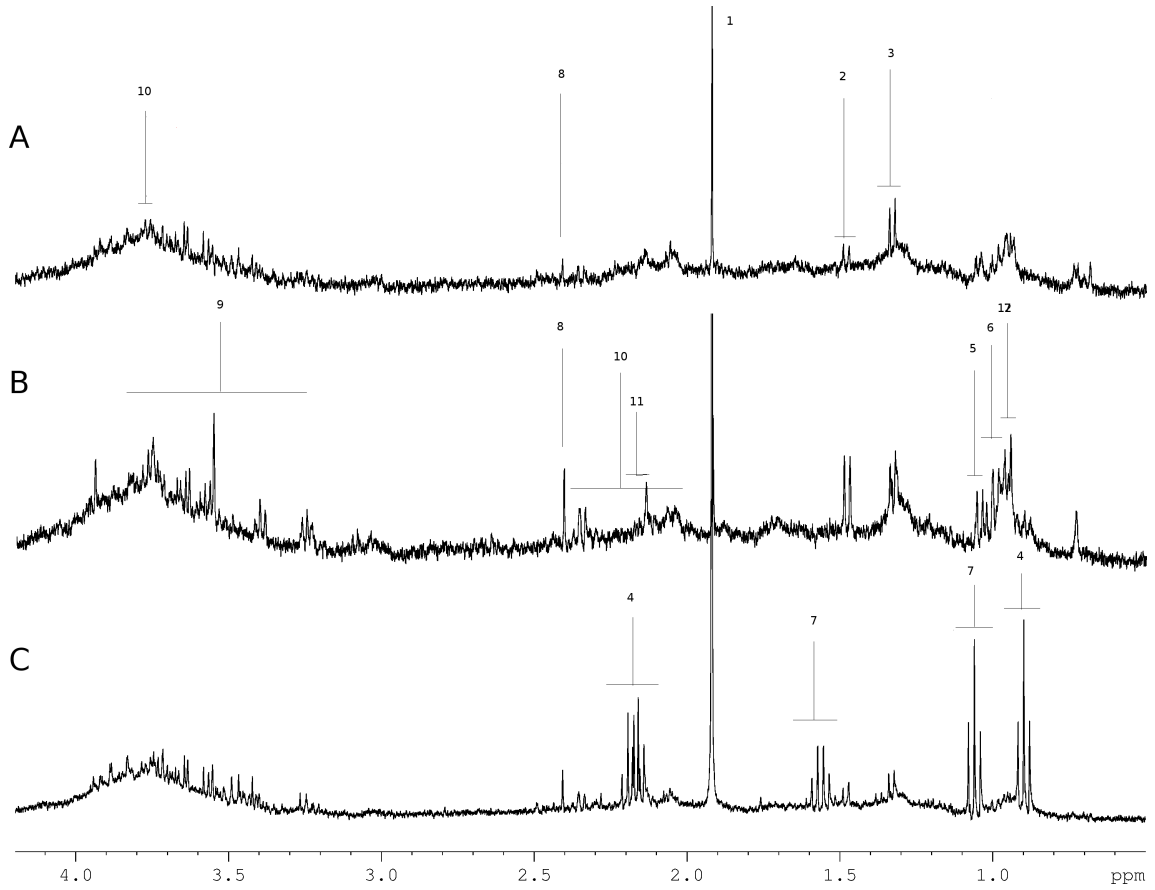


Figure 4.3: ***Representative ^{1}H NMR spectra (upfield region) from samples belonging to Cluster I (A), Cluster II (B), and Cluster III (C) as determined by the hierarchical clustering based on the relative percentage of genera in each sample.*** *Figure shows clear metabolic differences between spectra in each of the clusters. Spectra acquired on 400 MHz on a Bruker Avance spectrometer equipped with a 5 mm QNP probe using a zgesgp pulse sequence incorporating water suppression via excitation sculpting with gradients. Assignments: 1-acetate, 2-L-alanine, 3-lactate, 4-butyrate, 5-valine, 6-isoleucine, 7-propionate, 8-succinate, 9-glyceryl phosphate (?), 10-glutamate, 11-methionine, 12-leucine.*

4.3 Results and Discussion

Analysis of the composition of the gut microbial communities, associated with health and disease, and the gene content has been done extensively using sequencing [210–212]. The experimental evidence, however, shows that functional similarities are found between divergent communities and, conversely, distinct functional attributes can result from microbiota of similar composition [195, 213]. This highlights the importance of investigating the function of gut communities and not only the composition as only then we will be able to understand host-microbial interactions in depth and identify the compounds that mediate the resulting function.

Metabolomics allows simultaneous measurement of a large number of metabolites in a sample and the resulting data provide a fingerprint of the function of a given microbiota. Comparing NMR profiles of samples allows to cluster them according to the levels of metabolites in them. Such grouping of samples/microbial communities based on changes in metabolome allows to link them to changes in function over time or in response to a specific treatment or a condition [214].

^{1}H NMR spectroscopy was performed on buffered saline extracts from faecal samples numbered blindly from 1 to 20. It was hypothesised that there would be differences when comparing the metabolome of faeces from mice whose faecal microbiota were distinct. Initial examination of 1D ^{1}H NMR spectra from different clusters showed marked differences between metabolite levels (Figure 4.3 on the preceding page). PCA of spectra allowed comparison of all of the spectra simultaneously. The top panel in Figure 4.4 on the next page shows all the samples prepared in triplicate and recorded. Close clustering of the replicates indicates good reproducibility of the experiment and robustness of the method. The bottom panel shows the same samples coloured according to the cluster membership from the hierarchical cluster analysis of gut microbiomes: Cluster I, Cluster II and Cluster III. The samples are not grouped perfectly by cluster membership based on microbiota, which suggests that there are other discriminating factors such as contributions from the room group membership. To see which metabolites separate clusters from each other, further analysis involved a series of pairwise orthogonal partial least squares discriminant analysis (OPLS-DA) tests. Each comparison analysed spectra according to classes suggested by clustering according to the microbiota composition (Figure 4.2 on page 90) determined by

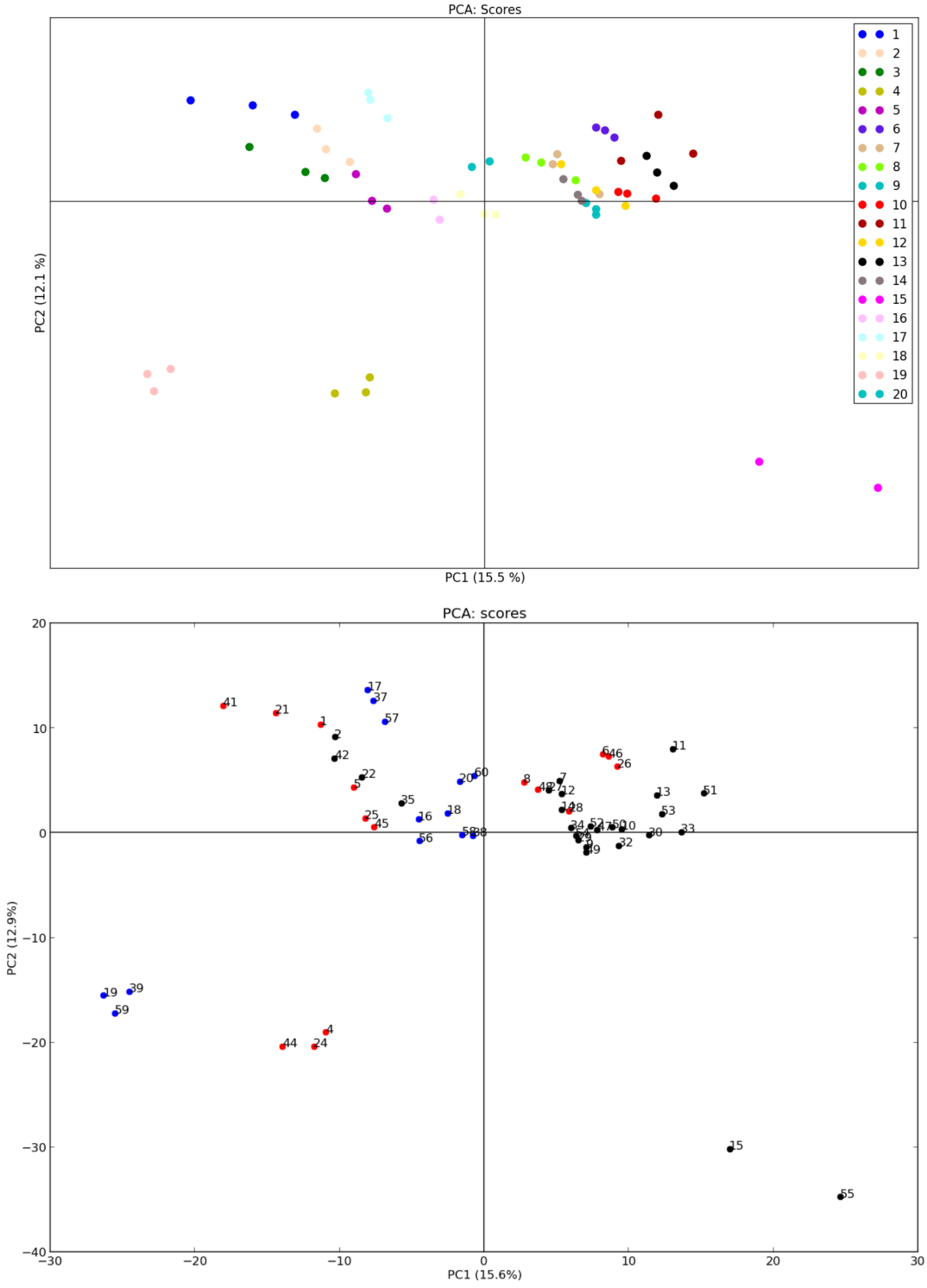


Figure 4.4: ***Scores plots from unsupervised PCA analysis*** *(spectra normalised using auto-scaling as used in OPLS-DA analysis). Samples are colour coded according to randomly assigned sample number (top) or cluster (bottom) from the hierarchical cluster analysis of gut microbiomes with black, blue and red representing, respectively, Cluster I, II, and III.*

16S sequencing.

Figure 4.5 on the next page presents scores plots which show a separation between clusters in the left panel: Cluster I vs Cluster II, Cluster I vs Cluster III, and Cluster II vs Cluster III. Cluster I and Cluster III show some degree of overlap, meaning that the metabolic differences between these two clusters were not as great as between other comparisons. This can be also seen from the Q^2, which for this comparison was 0.52 (Table 4.4 on page 104). Such a score indicates an effective separation of class members, however, it is lower than the other two comparisons and far from the ideal score of 1. The right side of the panel shows identified key drivers of the differences in the metabolomic data from the back-scaled loadings plots and assigned resonances with high variance and high weight, indicated by greater intensity and yellow/red color respectively (Figure 4.5 on the next page-right panels). Table 4.3 on page 98 lists all the peaks seen in the spectra along with tentative chemical shift assignments. Resonances were assigned based on multiplicities derived from JRES, the size of J-coupling and ^{1}H chemical shifts with reference to the *E. coli* metabolome database [58]. The analysis indicate that samples in Cluster II had higher levels of valine, alanine, isoleucine, phenylalanine and tyrosine as compared to faecal samples in Cluster I, samples in Cluster II had higher levels of butyrate, lactate and acetate as compared to Cluster I, whereas Cluster III had higher levels of acetate, butyrate, propionic acid and lower levels of valine and alanine as compared to samples in Cluster II. Notably, samples in Clusters I and II were distinguished by the greater abundance of a number of amino acids in the faecal metabolomes in Cluster II whereas the faecal metabolomes of mice from Cluster III were distinguished from those in Cluster I and II on the basis of short chain fatty acids which were more abundant for those in Cluster III.

4.3.1 PCA vs OPLS-DA

OPLS-DA was also carried out for comparisons dividing samples according to room occupancy or dominant phyla and the results are presented in Figure 4.6. Q^2 values obtained for each test were compared with a reference Q^2 value, obtained after repeating cross-validation with randomly generated class assignments and are presented in Table 4.4 on page 104. As shown, Q^2 values for the metabolomic data pairwise analysis performed when separated according to these clusters were > 0.50 which is considered a good model [107, 137]. As expected,

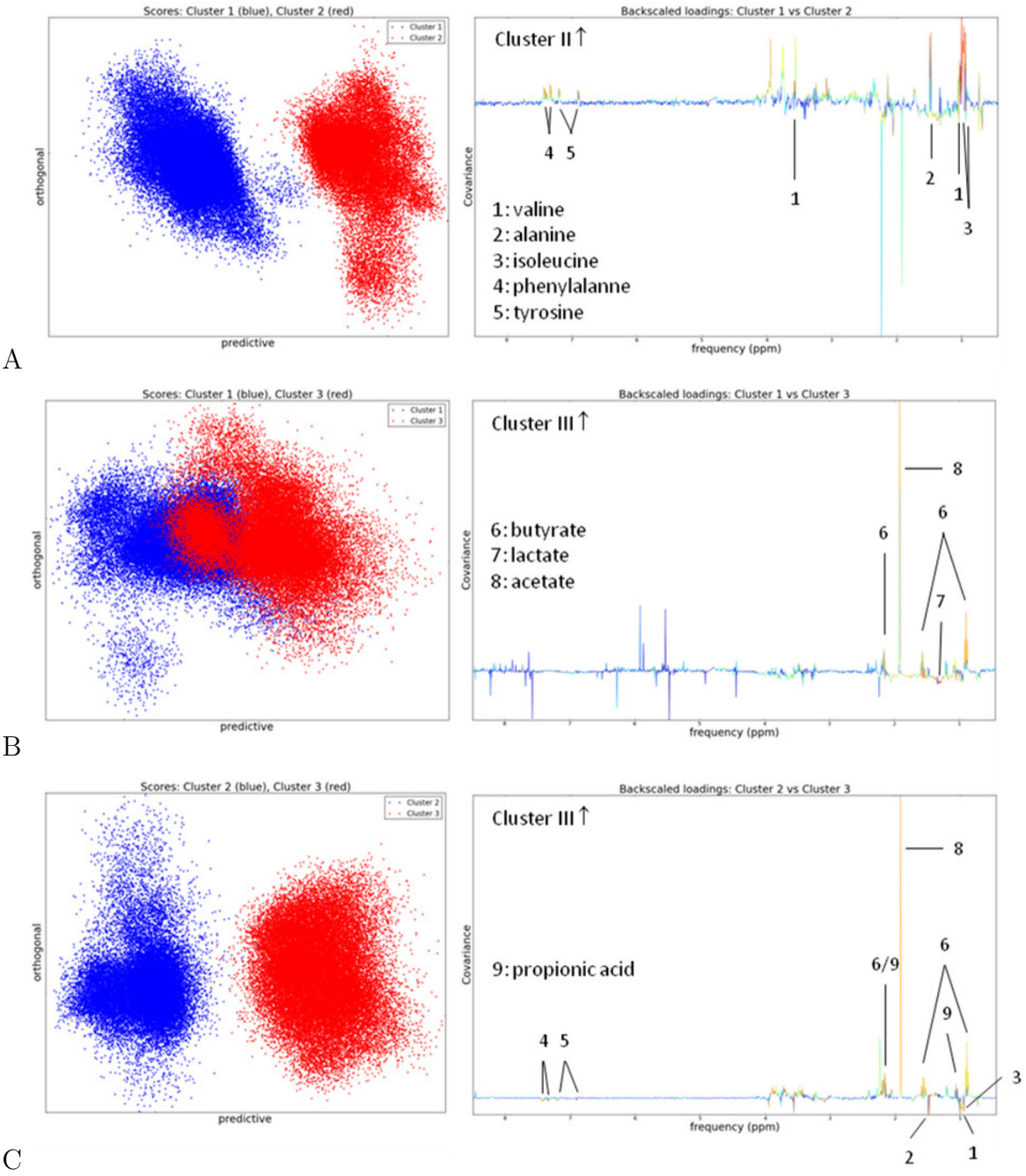
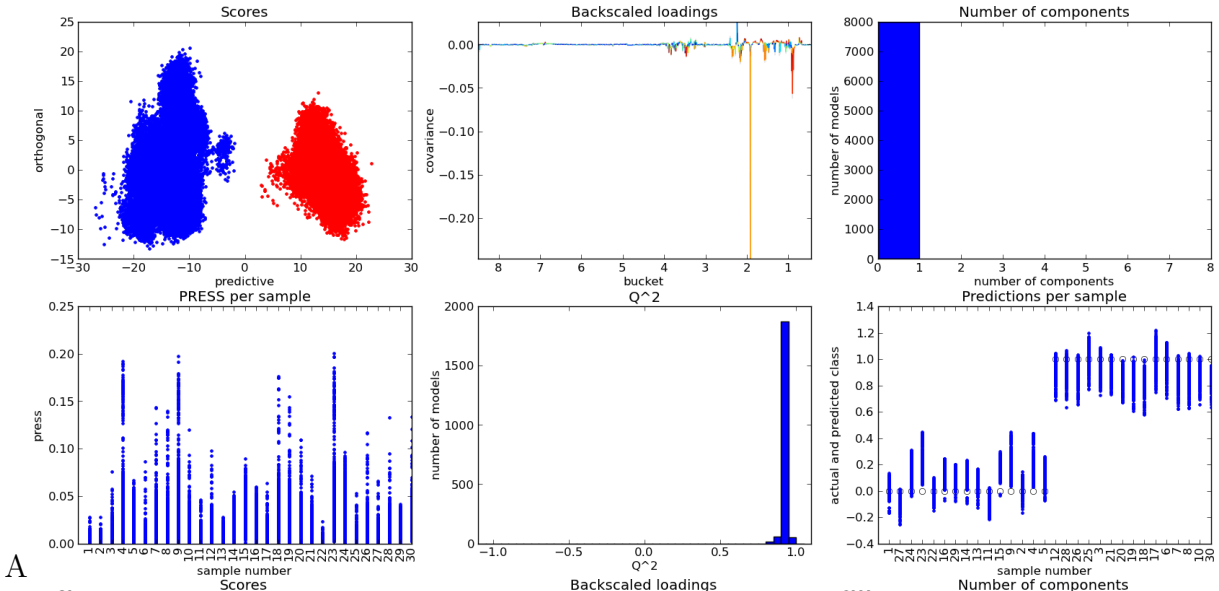


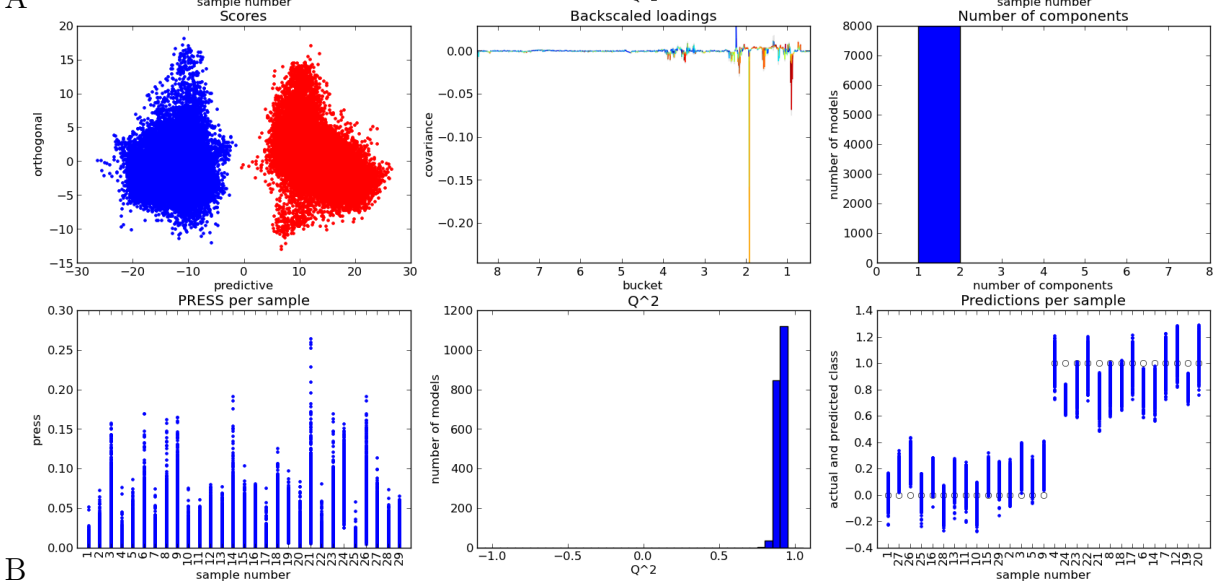
Figure 4.5: ***OPLS-DA scores plots (left panels) and back-scaled loadings plots (right panels)*** *with resonances with high variance and high weight highlighted in red for comparisons between the faecal metabolomes as clustered according to microbiome community. From the top: Cluster I vs Cluster II (A), Cluster I vs Cluster III (B), Cluster II vs Cluster III (C). Distinguishing metabolites that could be unambiguously assigned are annotated in each back-scaled loadings plot and the cluster with increased metabolite yield is indicated with an arrow.*

Table 4.3: *Tentative chemical shift assignment in ^1H NMR spectra of faecal samples.*

assignment	ppm	multiplicity	J coupling	peak
acetate	1.92	s		
butyrate	0.89	t	7.49 Hz	
	2.16	t	7.49 Hz	
	1.57	q	7.50 Hz	
proprionic	1.06	t	7.82 Hz	
	2.17	q		
lactate	1.34	d	7.06 Hz	
valine	1.05	d	7.14 Hz	
	1	d		
	3.7	d		
	2.25	m		small
isoleucine	0.94	t		
	1.02	d		
	3.66	d		
	1.97	m		small
	1.24	m		very small
	1.45	m		very small
alanine	1.49	d	7.37 Hz	
	3.79	q		
fatty acids	0.729	broad s		
dihydrothymine (?)	1.2	d	6.2 Hz	
	2.06	d		
	2.14	s		
	2.18	d		
	2.3	q or d		
	2.291	s		
succinate (?)	2.41	s		
	3.248	?		
glycine (?)	3.551	s		
lysine (?)	3.08	t		small
	3.74	t		
	1.91	m		small
	1.75	m		small
	1.5	m		very small
glycolic acid (?)	3.945	s		
phenylalanine	7.43	t	7.47 Hz	
	7.33	dd		
	3.98	dd		small
	3.1-3.3	m		small
tyrosine	6.9	d	8.2 Hz	
	7.2	d	8.2 Hz	
	3.93	dd		very small
	3.02-3.19	dd+dd		small

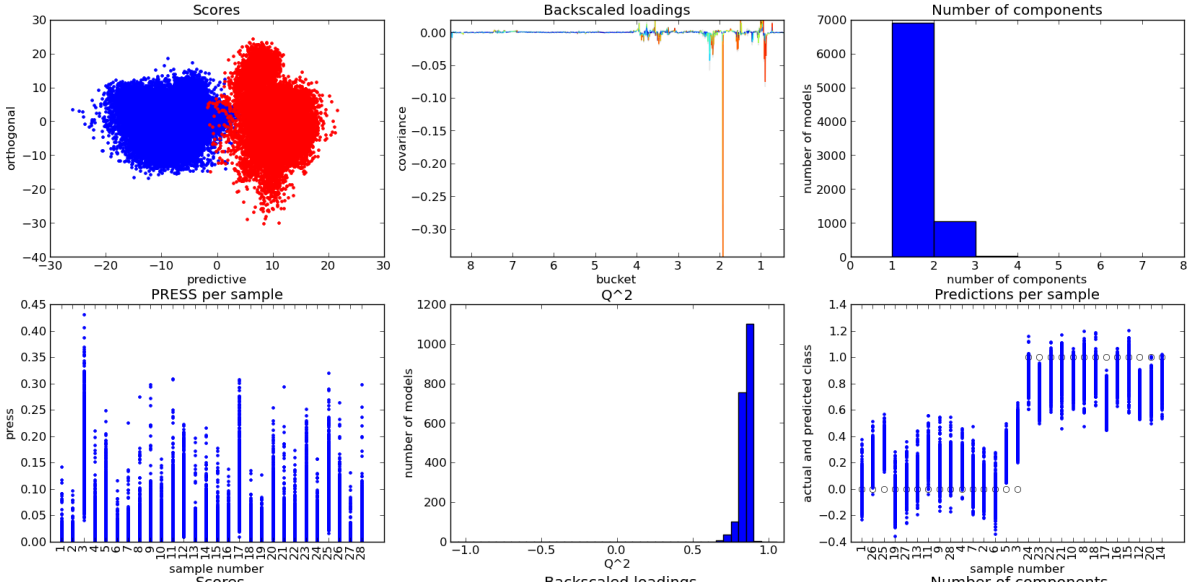

Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

A

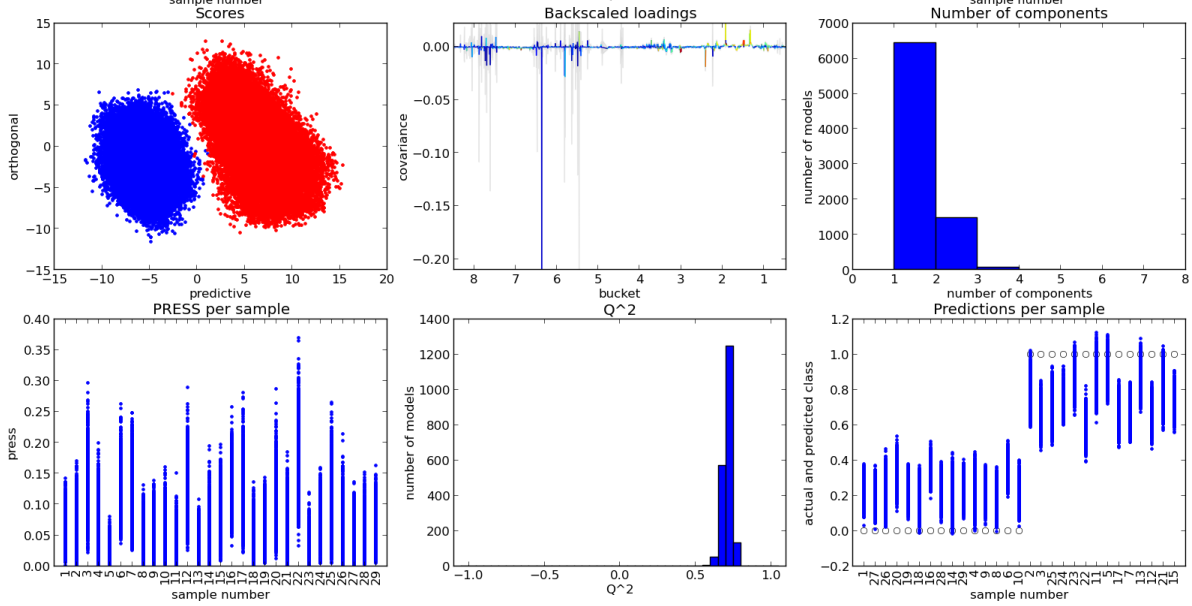

Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

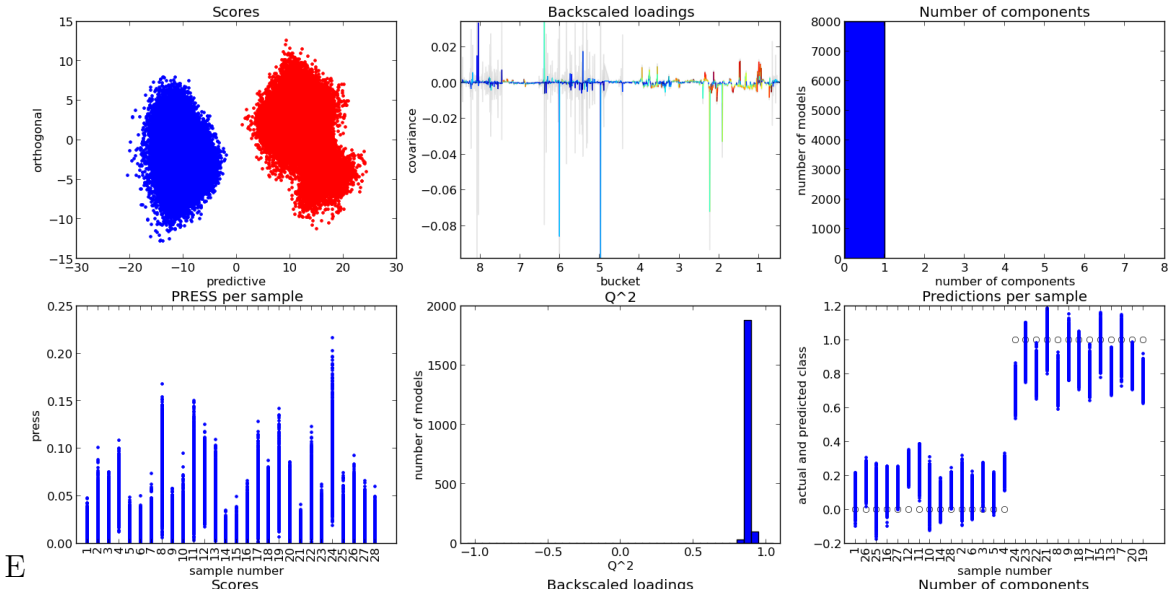
B

C

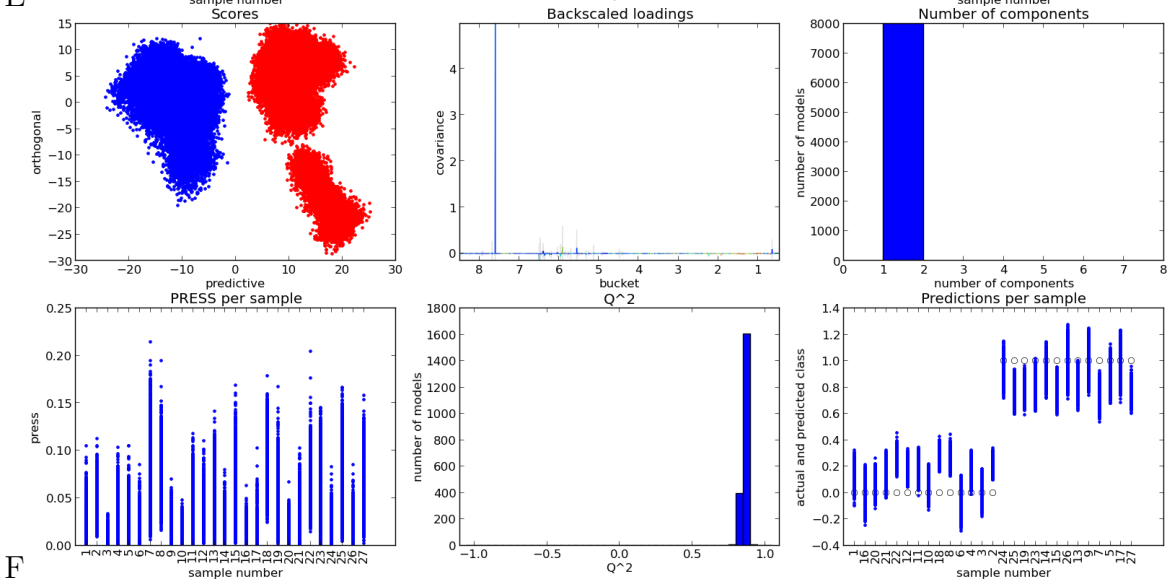

Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

D

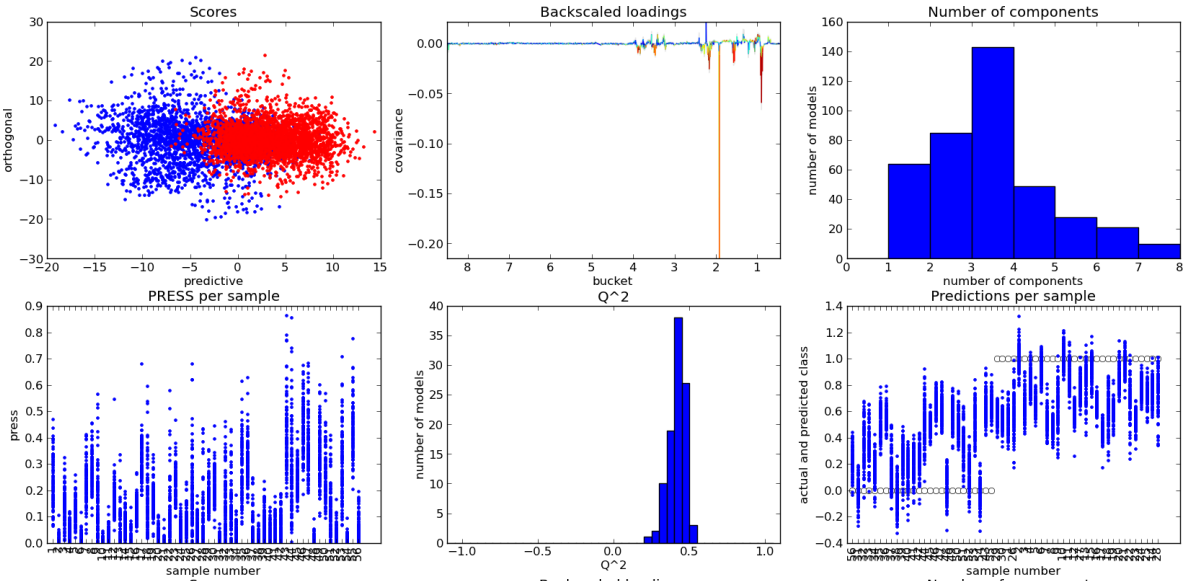

Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number


Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

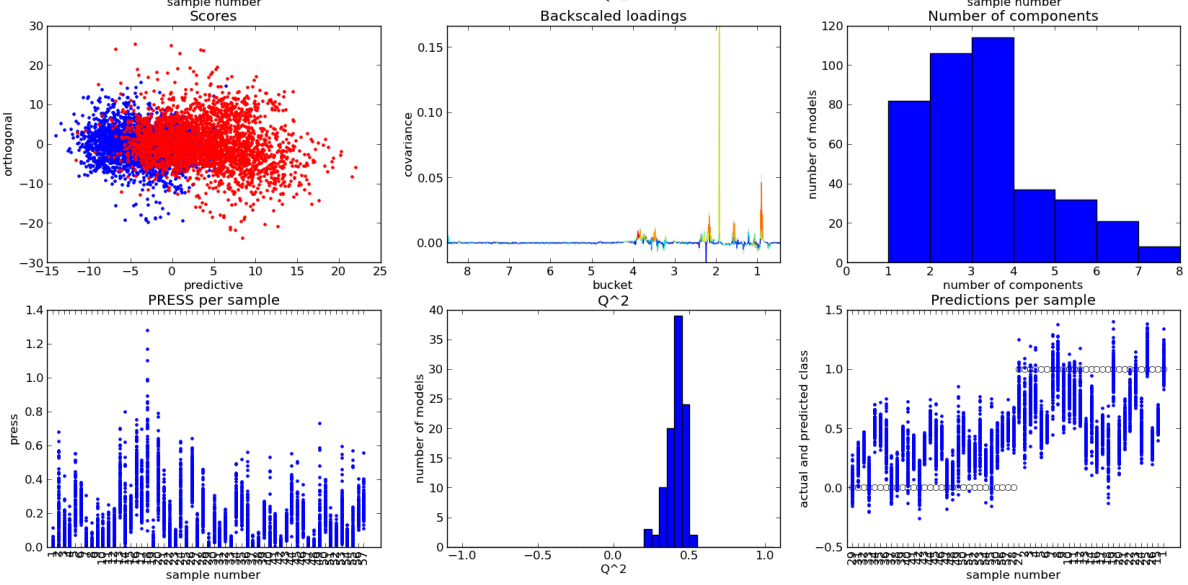
E


Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

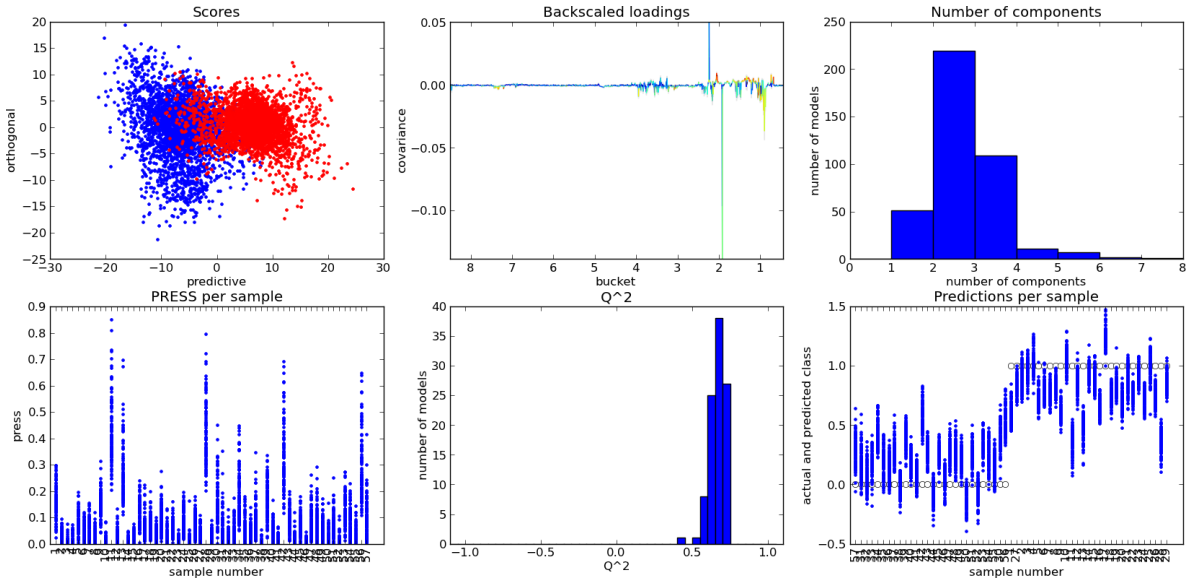
F


Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

G


Scores
orthogonal
predictive
Backscaled loadings
covariance
bucket
Number of components
number of models
number of components
PRESS per sample
press
sample number
Q^2
number of models
Q^2
Predictions per sample
actual and predicted class
sample number

H



I

Figure 4.6: ***Leave-one-out cross-validation (LOOCV) output files for the comparisons between room groups (A-F): room 1 vs room 2 (A), room 1 vs room 3 (B), room 1 vs room 4 (C), room 2 vs room 3 (D), room 2 vs room 4 (E), room 3 vs room 4 (F) and samples with high vs low percentage of Bacteroidetes (G), Firmicutes (H), Proteobacteria (I).*** *From top left to bottom right: scores plots showing separation between classes; back-scaled loadings plots showing* ^1H *resonance frequencies that discriminate between the two classes under comparison; histograms showing number of principal components used to separate the classes; prediction residual error sum of squares (PRESS) indicating low within-sample variation;* Q^2 *value histogram; comparison of model predicted for the sample classification as compared to the actual class. A double cross-validation was repeated 2000 times for comparisons A-F and 100 times for comparisons G-I.*

Table 4.4: ***Predictive Q^2 values for all models.*** *Q^2 values for models run with randomised class assignments are given in parentheses.*

Model	Q^2
Cluster I vs Cluster II	0.88 (-0.15)
Cluster I vs Cluster III	0.52 (-0.15)
Cluster II vs Cluster III	0.81 (-0.18)
Room 1 vs Room 2	0.93 (-0.14)
Room 1 vs Room 3	0.90 (-0.15)
Room 1 vs Room 4	0.85 (-0.15)
Room 2 vs Room 3	0.67 (-0.09)
Room 2 vs Room 4	0.80 (-0.12)
Room 3 vs Room 4	0.86 (-0.15)
High Bacteroidetes vs low Bacteroidetes	0.41 (-0.15)
High Firmicutes vs low Firmicutes	0.41 (-0.17)
High Proteobacteria vs low Proteobacteria	0.66 (-0.18)

clear metabolomic differences were observed between samples based on clusters defined by the composition of the bacteria present. Diet can influence microbial composition [215], however, in this study all mice received the same standard diet. One study [216] points out that the microbiome composition is unstable and can exhibit detectable changes within 24 h of changing the diet. Here and in other studies the selective mechanisms for divergence in bacterial composition remain largely unclear [195, 217] but could be driven by competition, inhibition and niche specialisation of bacteria or a combination of all these factors [214].

The relative contribution of Bacteroidetes, Firmicutes and Proteobacteria to each of the samples tested was also assessed using OPLS-DA. Here, Q^2 scores were lower (Table 4.4), which indicates poor models, i.e. the discriminating factor between the metabolite composition of samples was not (solely) the contribution from different phyla. OPLS-DA indicated significant differences in the metabolome of faeces from mice housed in different room groups with Q^2 scores all > 0.67, which indicates that housing alone may contribute to divergent gut metabolomes.

The best separation and the highest Q^2 was obtained for comparison between

rooms 1 and 2 and rooms 1 and 3, whereas the most overlap and the lowest Q^2 was obtained when comparing high and low Bacteroidetes content and low and high Firmicutes content. Nevertheless, the models separated the spectra and the backscaled loadings revealed similar patterns for comparisons between rooms 1 and 2, rooms 1 and 3, rooms 1 and 4, and also between high and low Bacteroidetes and Firmicutes content content. A similar pattern of discriminatory metabolite changes could be seen when comparing Cluster I and Cluster III (Figure 4.5 on page 97, B) and also Cluster II and Cluster III (Figure 4.5 on page 97, C). Other comparisons presented different pattern of altered metabolite levels again showing that the room membership is an additional factor in clustering seen in PCA (Figure 4.4 on page 95).

PCA of metabolomes extracted from faeces of mice hosted in different rooms found additional factors other than microbiota composition that drive the clustering pattern. Separation seen in PC1, which represents the main source of metabolome variation in the data set, in this case is not linked to OPLS-DA analysis, i.e. the discriminatory features in OPLS-DA are not the same as the features giving separation in PCA. This reveals that the room membership is an important factor differentiating between metabolome of mice hosted in different rooms. In the previous chapter, a study of *P. aeruginosa* strains isolated from sputum samples from cystic fibrosis patients, univariate patient data has been related to multivariate NMR data to show the functional effects of bacterial colonisation in a host. Here, two sets of multivariate data are presented-a hierarchical cluster analysis showing relative percentage of the associated genera in each sample and OPLS-DA showing differences in the metabolite composition between the samples. This allowed investigation of differences in host metabolome and align it with microbiota variation at the genus level.

4.4 Conclusion

Mouse models are commonly used in biomedical research and in order to avoid potentially confounding differences in genetic backgrounds, mice are taken from inbred populations to ensure homogeneity. When purchased for research, individual mice are considered identical and are expected to provide a uniform platform for study. The potential of the gut microbiota to influence the host in relation to health and a wide range of clinical syndromes is being increasingly recognised.

The variation in gastrointestinal microbiota composition is likely to be more significant when mice are moved between facilities, experience changes in diet, and are exposed to other animals or in experimental facilities, that are not highly controlled. Therefore, the divergence of mice gut microbiota identified in our metabolomic study, that was due to mice being housed in separate controlled units require further consideration and efforts are needed to ensure that mice used in research are indeed equivalent.

5

Combined systems approaches to understand pathogen-antibiotic interactions

Work described in this chapter has been published as:

Kozlowska, J., Vermeer, L.S., Rogers, G.B., Rehnnuma, N., Amos, S-B.T.A., Koller, G., McArthur, M., Bruce, K.D. & Mason, A.J. Combined systems approaches reveal highly plastic responses to antimicrobial peptide challenge in *Escherichia coli*. *PLoS Pathogens* 2014 (10)

5.1 Introduction

5.1.1 Antibiotic Resistance

Antimicrobial agents have been used for more than 70 years in the treatment of infectious diseases [218] and resistance development is a natural process. Most pharmaceutical antibiotics are derived from naturally occurring bacterial defences manufactured by fungi, bacteria, and algae against other fungi, bacteria, and algae. Natural antibiotics and the bacteria they are targeting have co-evolved over millions of years. During these battles bacteria have developed counter-measures in the form of genes necessary to produce and resist enemy antibiotics. Such arms races allowed bacteria to accumulate genes for bacterial resistance. However, the incorporation of antibiotics into animal feeds and the uncontrolled use of antibiotics in the clinical setting is likely to have led to loss of the balance and development of antimicrobial resistance [219–221]. Manufactured antibiotics have not been able to counteract bacterial adaptive responses and evolving new defences [220]. This has become an important concern in multiple healthcare contexts and is compounded by the scarcity of new therapeutic agents. The world is faced with drug-resistant bacterial infections, which cannot be treated by any of the current antibacterial options hence there is an urgent need for novel agents [222]. According to the Centers for Disease Control and Prevention (CDC) rates of infection due to both Gram-positive and Gram-negative pathogens are increasing [223]. Infectious diseases are known as one of the most life-threatening disabilities worldwide. Approximately 13 m deaths related to infectious diseases are reported each year [218]. Indeed, currently more people die in US hospitals due to meticillin-resistant *Staphylococcus aureus* (MRSA) infections than of HIV/AIDS and tuberculosis combined [224]. This is compounded by the emergence of panresistant Gram-negative bacteria such as *Acinetobacter* species, multidrug-resistant (MDR) *P. aeruginosa*, carbapenem resistant *Klebsiella* species as well as *E. coli* in most geographical areas [225]. According to the European Centre for Disease Prevention and Control (ECDC) infections due to these drug-resistant bacteria in the EU result in extra healthcare cost each year of at least €1500 m. Of particular concern is the increased percentage of *Klebsiella pneumoniae* (*K. pneumoniae*) resistant to carbapenems, current last-line antibiotics against these bacteria. In contrast, the percentage of (MRSA) isolates appears stable, and seems to be decreasing in some countries. However, MRSA remains a public health priority, as the percentage of MRSA continues to be high in several countries, especially in southern Europe.

It is therefore crucial to establish antibacterial strategies that will not only be able to treat drug resistant infections but, more importantly, foresee how resistance evolves. There is significant uncertainty, however, that such need will be met in the foreseeable future as, from 13 pharmaceutical leaders in the market of anti-infective drug discovery and development today, only a few big pharmaceutical companies (GlaxoSmithKline, AstraZeneca and Merck & Co) are still actively researching antibiotics. In the last 50 years, bacteria have developed resistance to every antibiotic within only few years of its release [226]. Since 2000, five new antibiotics have been launched, however, all of them were limited to treatment of Gram-positive infections [227]. It is therefore noteworthy that the overall number of compounds in development to treat Gram-negative infections, in particular, is very low. This reflects the challenge of developing treatments against Gram-negative bacteria, because of the presence of an outer membrane permeability barrier, multiple efflux pumps, and antibiotic- and target-modifying enzymes [228, 229]. There is ongoing effort to highlight the potentially disastrous outcomes of such situation, including the well-publicised annual report of the Chief Medical Officer of the United Kingdom in March 2013, which addressed the threat of antimicrobial resistance and called for antimicrobial resistance to be put on the national risk register [230].

Mechanisms of antimicrobial resistance

Interactions between bacteria and their environment, and host in the case of pathogens or commensals, provide a selection pressure for bacterial adaptation that supports increased survival [231]. The network of interconnected and tightly controlled stress responses not only protects bacteria, but also affects their antimicrobial susceptibility via modifications to the cell at a cellular, metabolic and genetic level. The common characteristic of many stress responses is activation and upregulation of error-prone DNA polymerases and simultaneous downregulation of enzymes responsible for DNA damage repair and movement of transposons (mobile genetic elements) leading to point mutations and recombination [231].

The SOS response to antibiotic treatment has been demonstrated in numerous studies to induce toxin production [232–235], propagation of virulence factors and antibiotic resistance genes. In Gram-negative bacteria, resistance determinants are carried in the mobile gene casettes (integrons), which are controlled by the SOS response and LexA-RecA regulon in *E. coli* and *Vibrio cholerae* (*V.*

cholerae) [236]. Moreover, the SOS response stimulates high mutation rates, which might relieve the stress but also facilitate acquisition of antibiotic resistance genes.

Exposure to different classes of antibiotics can induce resistance development via different mechanisms. Penicillin and other β-lactam antibiotics induce the SOS response by interfering with cell wall synthesis [237], therefore cells become temporarily resistant by inhibiting cell division. Resistance upon exposure to β-lactam antibiotics can be also a result of a lack of autolytic enzymes, secretion of β-lactamase, activation of efflux pumps and also modifications to its target-penicillin binding proteins (PBPs) due to accumulation of (p)ppGpp and subsequent induction of the stringent stress response [238, 239]. Antibiotics of the quinolone class have been demonstrated to be mutagenic [240]. Their primary target is topsoisomerase IV and gyrase and mutations to the genes encoding these enzymes could be responsible for the rise of antimicrobial resistance, therefore exposure to them could induce resistance as demonstrated by Cirz *et al.* [241]. Furthemore, Beaber *et al.* [242] showed that *V. cholerae* exposed to quinolone, acquired resistance to antibiotics not by mutagenesis, but via movement of integrating conjugative elements (ICEs). ICEs are one of the three main types of selftransmissable mobile genetic elements and key mediators of horizontal gene flow in bacteria. Similarly to plasmids, they transfer during the process of conjugation, however, they are able to integrate into the host chromosome, replicate and excise to transfer again, which is typical of bacteriophages [243]. Sublethal doses of streptomycin, on the other hand, produce mutations in bacteria via an SOS-independent fashion, mainly by interfering with ribosomal translation [244]. Aminoglycoside resistance in strains of *P. aeruginosa* is a result of exposure to the treatment and ribosome disruption, which activates the MexXY-Opr multidrug efflux system [245]. In a recent study by Dalebroux [246] it was shown that *Salmonella typhimurium* (*S. typhimurium*) becomes resistant to polymixin B and the mammalian antimicrobial peptide C18G after exposure by making modifications to the component of LPS, Lipid A, and also glycerophospholipids (GPLs) in the inner leaflet of the OM. It has also been shown that exposure of *E. coli* to sublethal doses of a range of antibiotics regardless of their mechanism of action resulted in multidrug resistance [247] and one of the proposed mechanisms was increased mutation rate due to raised levels of ROS [248].

However, antimicrobial exposure is not the only way in which bacteria can

acquire antimicrobial resistance. Research shows that bacterial natural environment and the stress responses that are triggered in response to environmental changes are sufficient to diminish innate antimicrobial susceptibility and generate resistant mutants even without exposure to antimicrobials due to e.g. oxidative stress [245, 249], nitrosative stress, nutrient limitation, membrane damage, heat and ribosomal stress [245] or selective pressure exerted by the host [46]. This can be an indirect effect, merely due to growth inhibition or bacteria entering a dormant state [237, 250], or a direct result of inducing a myriad of bacterial stress responses. They may result in changes promoting antibacterial resistance such as activation of efflux pumps [245], remodelling of antimicrobial targets [238, 251], induction of resistant growth modes such as biofilms [252] and generation of mutations leading to increased resistance [35].

The ability to suppress bacterial stress responses, mainly the SOS response, has become of interest and seems a likely goal for treatment of some of the bacterial infections. There is also evidence that diversifying antibiotic prescription patterns can help increase bacterial susceptibility to certain antimicrobials [253].

5.1.2 Antimicrobial peptides and innate immunity

Faced with multidrug antimicrobial resistance researchers turned towards agents with broad-spectrum activity and mechanisms of action with proposed lower propensity for resistance development: antimicrobial peptides (AMPs). The reason to study AMPs is twofold. First, both bacteria and AMPs are ancient; they have evolved together and studying bacterial responses to a wide range of AMPs can provide insight into the inherent ability of bacteria to respond to challenges and the adaptive mechanisms available to overcome them. The second reason is that AMPs have prospects to be developed into anti-infective therapeutic agents as antimicrobials, adjuvants for synergistic effect or as immunomodulatory and/or endotoxin-neutralising compounds [254–258].

AMPs are polypeptide antimicrobial substances with fewer than 100 amino acid residues, that are secreted from various cells and tissues of many invertebrates, plants or animals but also archaea, eubacteria, protista. They have shown direct activity against Gram-negative and Gram-positive bacteria, fungi, viruses and protozoa [25, 259].

AMPs represent a large and diverse group of molecules that can be classified

according to their amino acid composition or structure. It is the class of cationic antimicrobial peptides that offers potential in treatment of microbial infections, but also as antivirals, immunomodulators and antitumoral drugs [260]. The disadvantage of anionic antimicrobial peptides is that they often require cationic moieties such as synergistic cationic antimicrobial peptides or cations such as zinc (Zn^{+2}) to act as cationic linkage between the anionic antimicrobial peptide and the anionic microbial cell membrane [261]. As of October 2013 over 2000 different antimicrobial peptides and synthetically created peptides are known and their number is constantly increasing [262]. Several databases exist that catalogue antimicrobial peptides on the basis of the source organism, biological activities and specific peptide features. Currently the largest database is maintained by the Wang group http://aps.unmc.edu/AP/main.php [262].

There is little structure and sequence homology between peptides recovered from different species, even those that are closely related [254]. Despite such variability, there are a few common features including: small size (12-50 amino acids long), hydrophobicity and nominal cationic charge ranging from +2 to +9 [263]. Some antimicrobial peptides are secreted constitutively, while others are secreted in response to invading microbes and the presence of bacterial signalling molecules e.g. LPS and lipoteichoic acid (LTA) and bacteria [25].

Despite being exposed to AMPs for millennia, bacteria have not yet developed full resistance to broad spectrum AMPs. Nevertheless, AMPs can still represent a trigger for resistance development via alteration of cell surface charge and electrostatic repulsion of AMPs, activation of efflux pumps to extrude AMPs, production of proteases to cleave AMPs or trapping proteins [264, 265]. Moreover, unlike antibiotic resistance genes which are located on laterally transferable elements such as plasmids or transposons, most genes involved in resistance to AMPs are on the chromosome, close to housekeeping genes, and are considered as the integral part of the genome of some bacteria [266]. Knowledge of the molecular basis of bacterial AMPs resistance may therefore provide new targets for antimicrobial therapeutics.

AMPs are currently considered to act mainly at the level of cytoplasmic membrane and outer membrane as pore-forming molecules, but have also been reported to disrupt nucleic acids, affect protein synthesis, inhibit enzymes and sequester nutrients essential for the bacteria [25]. A number of leading review

articles [267–270] suggest that they may also have other immunomodulatory activities important for the anti-infective host defence. Immunomodulatory properties of AMPs reported in recent studies include the modulation of the levels of proinflammatory cytokines and chemokines by altering gene expression of host cells, activation of host immune cells (monocytes, macrophages, neutrophils) and ROS and RNS, stimulation of angiogenesis and wound healing and recruitment of leukocytes to the site of infection. These activities are not directly antimicrobial, nevertheless, they result in protection of the host against infection and allow selective control of the infection and immune responses [258, 271, 272]. They play important roles in preventing infection, but also in the clearance of infection and are thus often referred to as host defence peptides in contrast with peptides acting directly on bacteria and having a microbicidal effect.

Structural classes of AMPs and determinants of antimicrobial activity

The multidimensional properties of AMPs described above and a broad-spectrum antimicrobial activity are possible due to tremendous structural, and thus functional, diversity of those molecules. At least four structural groups have been proposed for AMPs according to their conformations: amphipathic α-helices, β-sheet structures stabilised by disulfide bridges, extended structures with one or more amino acids being predominant and loop peptides with only one disulfide bridge. Despite the differences at the primary and secondary structural level, most cationic antimicrobial peptides are amphipathic, i.e. they have a hydrophobic region, but also a positively charged hydrophilic structural domain and therefore can exist at the interference between polar and non-polar environments. The positive charge allows peptides to interact with negatively charged targets including DNA, RNA etc. and also bacterial membranes, as the initial interactions are electrostatically driven due to the negatively charged membranes, whereas amphipathicity facilitates binding and insertion of the peptide into the hydrophobic core of the lipid membranes [273].

Several studies have demonstrated the importance of an α-helical structural domain, showing that the increase in the amount of α-helical structure is related to enhanced antimicrobial activity [274, 275] or broadening of the spectrum of activity [276]. In contrast, more disordered peptides that have the ability to disorder the lipids in their vicinity perform better than peptides that remain structured (manuscript in preparation). This was also observed in a molecular

dynamics simulation study where α-helical conformation was not required for pore formation and peptides were not highly structured. Indeed, experimental studies involving stabilised α-helical peptides have shown reduced antimicrobial activity [277] and that partial unfolding might facilitate pore formation [278]. Consequently, despite being very intensively studied experimentally, relatively little is known regarding the mechanism of pore formation or the structure of the pore itself. It is commonly accepted that many antimicrobial peptides form toroidal pores and this molecular dynamics simulation study demonstrated that a threshold concentration is required to create enough stress in the membrane and that pore formation is facilitated by aggregation of the peptide on the membrane [278].

Other features known to be important determinants of antimicrobial activity are amphipathicity, hydrophobicity and charge [279]. The parameters are intercorrelated and modification to one can be compensated by changes in others, hence the optimal antimicrobial efficiency is a result of balanced coordination of all those factors [280, 281].

The lipid composition of the cell membrane of multicellular animals plays an important role in the selectivity of AMPs for their target. First the negative charge on the bacterial surface allows preferential binding of polycationic antimicrobial agents such as AMPs. In the mammalian cell the bilayer is asymmetric-the outer leaflet is composed of mainly phosphatidylcholine (PC) and sphingomyelin (SM) with zwitterionic head groups and the cytoplasmic surface is composed of the negatively charged lipids such as phosphatidylserine (PS), phosphatidylinositol (PI) and zwitterionic phosphatidylethanolamine (PE). The cholesterol present in mammalian membranes reduces the activity of AMPs, by stabilising the lipid bilayer and interacting with the peptide [254, 282]. Although the asymmetry of the *E. coli* inner membrane is unknown it is principally composed of zwitterionic PE (80%) and at least 15% of negatively charged phospholipids such as phosphatidylglycerol (PG) or cardiolipin (CL) allowing formation of strong electrostatic interactions with polycationic antimicrobial agents [254, 283].

Modes of action of AMPs

Despite the vast number of known AMPs, their exact mode of action remains unclear or controversial in many cases. One of the reasons for this is that con-

centrations of AMPs used in some studies exceed the concentration needed to see a bacterial response. AMPs, as any other antibiotic, require a minimum concentration to have a growth inhibiting effect against a given bacterial strain, which is often expressed in terms of minimum inhibitory concentration (MIC). The concentrations above the MIC usually have an immediate macroscopic effect such as loss of viability or membrane lysis. It is not clear, however, what are the molecular mechanisms that induce inhibition when crossing the threshold dividing proliferating bacterial populations from nonproliferating ones. Possibilities include enzyme inhibition leading to disruption of bacterial metabolism or AMPs acting in a cooperative manner due to the increased concentration [278, 284]. A study by Huang *el al.* [284] exposed alterations in the behaviour of AMPs when certain molecular threshold concentrations were reached including conformational changes and differences in the association state as well as changes in the membrane topology such as poration. Such events associated with reaching the threshold concentration are potentially bactericidal, however, so far they were observed only in model lipid bilayers, with some AMPs reaching full membrane saturation, and not in *in vivo* studies. Those considerations apply to the AMPs with a membrane disruptive mechanism of action. For non-disruptive AMPs, where the primary aim is cytoplasmic invasion, parameters such as the partition constant, aqueous phase and membrane-bound peptide concentration do not correlate directly with the disruptive action of the peptide in the cytoplasm [285]. Since AMPs are cationic, they likely cause partial or complete neutralisation of the bacterial membrane upon binding and indeed a number of AMPs display membrane charge neutralising properties with other concomitant threshold effects [285–287] implying that membrane charge neutralisation and membrane saturation could constitute the main killing strategy for some AMPs. Another problem is that quantitative structure-activity relationship (QSAR) studies look at isolated processes in much simplified systems and the findings may not be reflected in *in vivo* studies [288]. Modifications to an AMP to promote one activity may inhibit another. This emphasises the need for a novel high-throughput method that would allow to probe the activity of AMPs in the whole bacterial cell.

Until recently, the existing dogma was that most cationic AMPs kill through membrane disruption or pore formation [289]. The initial interaction between peptide and the OM is thought to occur due to electrostatic attraction between cationic peptides and either negatively charged LPS in Gram-negative bacteria or LTA in Gram-positive bacteria (Figure 5.1 on page 118). Lipopolysacchar-

ride (LPS), also known as endotoxin, is the main constituent of the OM found in Gram-negative bacteria. It is a highly conserved, amphiphilic lipid that consists of three regions: O-polysaccharide (the O- or somatic-antigen), the core polysaccharide known as 2-keto-3-deoxyoctonic acid (KDO) and lipid A. Lipid A is anchored in the OM [290]. AMPs distort the integrity of the OM by breaking the non-covalent associations between LPS and divalent cations such as calcium (Ca^{2+}) and magnesium (Mg^{2+}), facilitating insertion of the peptide into the membrane by a process known as self-promoted uptake [279].

Different killing mechanisms have been proposed that can occur at the membrane including: fatal depolarisation of the membrane, pore formation by various models and subsequent cellular leakage, formation of transient channels, micellarisation of the membrane or activation of deadly processes such as production of hydrolases that degrade the cell wall, thereby distorting lipid distribution between leaflets of the lipid bilayer [279]. Experimentally, it has not been possible to directly determine the number of peptides within the pore. The commonly cited number of 4-7 peptides [291, 292] has been indirectly inferred from a combination of neutron scattering data and oriented circular dichroism (OCD) measurements. There is considerable evidence, however, that some peptides target intracellular components upon entry to the interior of the cell rather than the membrane itself [289]. Having tools that enable to study how AMPs achieve their killing effect and in what conditions is of particular importance as this can potentially lead to better understanding of the pathologies and inform the development of AMP-based therapeutic agents.

In this study structurally and physically related AMPs were applied: magainin 2, pleurocidin and buforin II. Also, based on understanding of these three naturally occurring peptides a range of D-peptides have been designed comprising D-amino acids only in an attempt to circumvent the effect of proteases secreted by target pathogens [293], and incorporate structural features, including high cationicity and propensity for adopting α-helix rich conformation [294–296] and proline kink. Incorporation of proline on the hydrophilic face of the amphipathic α-helical AMPs increases the conformational flexibility and has been linked to improvement of antibacterial activity and reduction of the hemolytic effect [297–299]. Each of those peptides have been described to operate by different killing mechanism while the opportunity was taken to understand the mechanism of action of D-LAK120-AP13 for the first time. As described in the previous chapter,

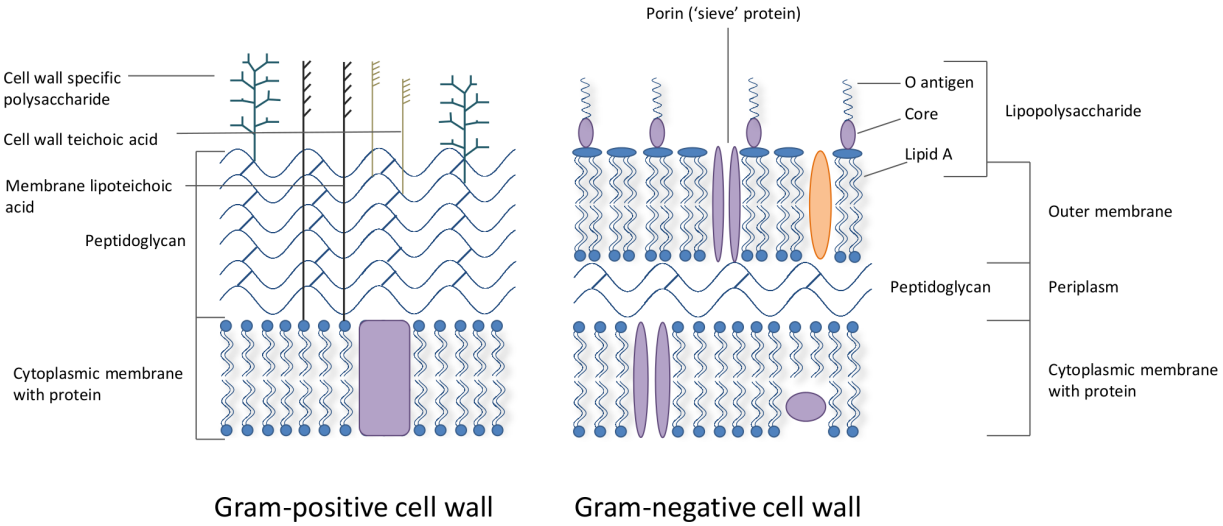


Figure 5.1: ***The structure of the cell wall of Gram-positive and Gram-negative bacteria.*** *Modified from Fierke Research Group (http://www.umich.edu/~caflab/lpxc2.htm).*

bacteria are able to harness a vast range of stress responses in order to overcome or adapt to the threat posed by AMPs and choosing peptides with various mechanisms of action was an attempt to probe the variety of these responses.

Considering their bactericidal strategies, peptides can be generally categorised as membrane disruptive and membrane non-disruptive, with the former being the dominant mechanism [25]. Magainin 2 [300], a cationic α-helical antimicrobial peptide (Figure 5.2 on the next page, a) isolated from the skin of African Clawed frog *Xenopus laevis*, is considered to have a membrane disrupting activity towards both Gram-positive and Gram-negative bacteria. It is thought to bind preferentially to negatively charged phospholipids present in bacterial membranes with the formation of a dynamic peptide-lipid pore and subsequent cell permeabilisation by a mechanism known as the toroidal pore model where the lipid bends back on itself. The detailed molecular mechanism of this model has been described [291]. Apart from the toroidal pore model there are at least three other proposed models by which peptides are able to aggregate and/or re-orient in the membrane and disrupt its integrity: the barrel-stave [259], the aggregate [301] and the carpet model [302].

Membrane non-disruptive peptides have alternative mechanisms of action that involve the interaction of the peptide with the membrane which does not necessarily result in membrane disruption; their main targets are polyions such as

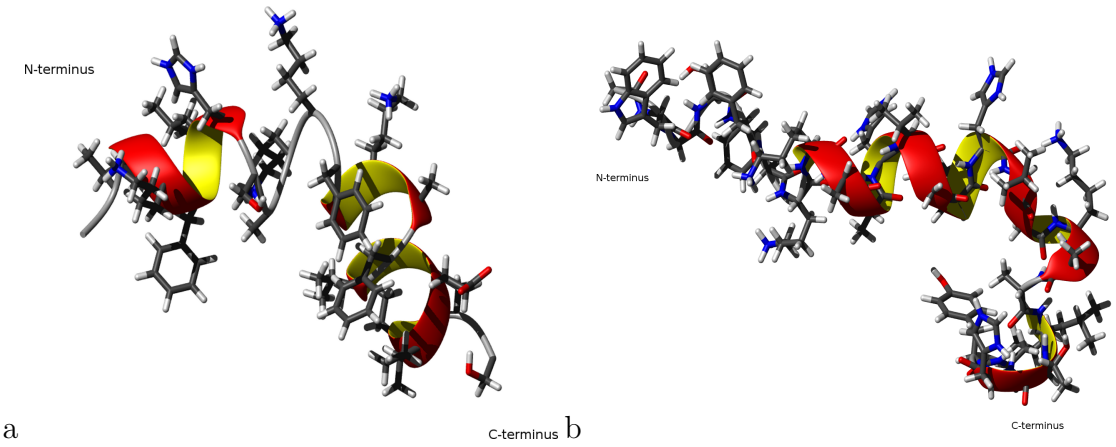


Figure 5.2: ***Structures of SDS-bound magainin and pleurocidin*** *solved in our laboratory showing side chain orientation and backbone topology. Images created with MOLMOL software and rendered with POVRay. The typical conformers from the ensembles closest to the average structure are shown. PDB entries for magainin 2-2LSA and pleurocidin-2LS9.*

DNA or RNA. Once they penetrate the bacteria, they can potentially kill by interfering with bacterial metabolic processes [259, 289, 303]. An example of a well studied membrane non-disruptive peptide is buforin II, a 21-amino acid long cationic peptide derived from the Asian toad *Bufo gargarizans*, which is known to form an extended helical structure [304]. It was shown to have the ability to translocate through bacterial membranes and bind to DNA in a cooperative manner with high affinity forming peptide-DNA complex [64, 304]. The proline kink in buforin II is known to be crucial for enabling translocation into the *E. coli* cytosol [305]. However Buforin II has only barely detectable antibacterial activity against planktonic *E. coli* cultures [64].

Pleurocidin is a naturally occurring peptide showing antimicrobial activity against Gram-negative as well as Gram-positive bacteria. It was isolated from the skin and intestine of the winter flounder *Pleuronectes americanus* [306]. This 25-mer peptide has an amphipathic helical structure very similar to that of magainin 2 [306] (Table 5.1 on the following page). The study by Yoshida *et al.* [307] showed that pleurocidin, similarly to magainin, was able to translocate across the model membrane and to cause dye-leakage, however, when lower levels of peptide were used (at MIC_{50}) a pleurocidin analogue was demonstrated to act on RNA and protein synthesis by inhibiting histidine incorporation without concomitant membrane permeabilisation [308] suggesting that the peptide has multiple modes of activity. Pleurocidin was also shown to associate strongly

Table 5.1: ***Comparison of physical and biological features of peptides used in this study.*** *Hydrophobicity (H) and mean hydrophobic moment (μH) are shown according to the Eisenberg scale [1] and were calculated using the HydroMCalc Java applet made available by Alex Tossi [2]. Proline residues are underlined. Peptides in italics are D-amino acids. *Mean hydrophobic moment assuming formation of ideal α-helix.*

Peptide	Sequence	Charge	(H)	(μH)*	MIC_{50}(μM) *E. coli* NCTC 9001	MIC_{50}(μM) *P. aeruginosa* PAO1
Buforin II	TRSSRAGLQFPVGRVHRLLRK	7	-0.37	0.30	>26.30	>26.30
Magainin 2	GIGKFLHSAKKFGKAFVGEIMNS	4	-0.03	0.28	6.86±1.58	10.60±2.21
D-LAK120-AP13	*KKLALALAKKWLPLAKKLALALAKK*	9	-0.10	0.21	1.08±0.62	1.22±0.19
Pleurocidin	GWGSFFKKAAHVGKHVGKAALTHYL	5	-0.02	0.22	0.66±0.22	1.65±0.87

with anionic lipid head groups, where it forms ion channels in planar lipid bilayers most probably by formation of toroidal pore [273]. Proton decoupled ^{15}N solid state NMR spectroscopy showed that at neutral pH pleurocidin is oriented parallel to the membrane surface, whereas ^{2}H solid state NMR revealed that it disrupts membranes containing anionic lipids (phosphatidylglycerol (PG)) more effectively than zwitterionic (phosphatidylethanolamine (PE)) [309]. Moreover, intrinsic fluorescence measurements showed that pleurocidin insertion into lipid membranes occurs exclusively in the presence of anionic lipids as does strong dye leakage and increased peptide translocation across the membrane [307, 309]. In the presence of sodium dodecyl sulfate (SDS) micelles the peptide adapts α-helical conformation (Figure 5.2 on the previous page, b).

A number of AMPs while having membrane disruptive activity, have been shown to operate through intracellular modes of killing, such as binding nucleic acids [304, 310] or inhibiting nucleic acid or protein synthesis [308, 311, 312] and the exact modes of action of AMPs remains an area of controversy [256, 313].

This again highlights the need for a new method which would allow to investigate the mechanisms of action of AMPs using a holistic approach.

5.1.3 Systems biology: omics metodologies

The understanding of how AMPs function is therefore far from complete. Attempts to optimise AMP potency in the laboratory, that focus on only one possible bactericidal mechanism, ignore the possibilities offered by taking a holistic approach that can reveal the true source(s) of bactericidal potency along with a better understanding of bacterial counter-measures. The purpose of combining metabolomics and transcriptomics is to obtain a more comprehensive insight on system level adjustments in bacteria under stress from AMPs. The full power of -omics based research tools has yet to be brought to bear in antibiotic research [314] and the detailed mechanistic analysis of host-pathogen systems, encompassing all aspects of such a complex interaction is still in its infancy. Examples of previous studies include those that focus on Gram-positive bacteria (*Bacillus subtilis* [315], *Staphylococcus aureus* [316], *Streptococcus pneumoniae* [317]) which have demonstrated the existence of complex regulatory patterns in which several signal transduction pathways were induced. In our laboratory the focus is on trying to understand the relative difference in antibacterial potency of structurally related AMPs to Gram-negative bacteria such as *Escherichia coli* and *Pseudomonas aeruginosa* [64, 318–320]. Since these peptides act at widely differing effective concentrations it was hypothesised that studying their effects at sub-lethal concentrations would provide a detailed overview of the mechanisms of action of each AMP. Therefore, a method has been devised that could efficiently identify conditions where bacteria responded to AMP challenge without introducing possible, non-specific complications that might result from large scale cell death. ^{1}H high resolution magic angle spinning (HR-MAS) NMR was used to identify the lowest AMP concentration that elicited a response from metabolically active, challenged bacteria. A robust, cross-validated, multivariate analysis identified metabolites whose levels were altered in response to AMP challenge. These were used to classify the AMP according to the elicited response whilst providing a first indication of whether *E. coli* responded in a generic or specific manner to AMP challenge. Having identified sub-lethal conditions where a response was confirmed, electron microscopy and transcript profile analyses enabled a detailed description of the *E. coli* response to AMP challenge.

5.2 Materials and methods

5.2.1 Materials

Peptides (Table 5.1 on page 120) pleurocidin, magainin II, buforin 2 were purchased from Pepceuticals Ltd (Nottingham, UK) as desalted grade and further HPLC purified using water/acetonitrile gradients using a Waters SymmetryPrep C8, 7 m, 19300 mm column. D-LAK120-AP13 was synthesised by V. Abbate (King's College London). *Escherichia coli* (*E. coli*) NCTC 9001 was a gift from K. D. Bruce (King's College London). All other reagents were analytical grade or better.

5.2.2 Bacterial culture and challenge

Cultures of *Escherichia coli* NCTC 9001, a strain isolated from a patient with cystitis, were grown overnight in Mueller-Hinton broth (MH) at 37 °C . Once the OD_{620} reached ≈1.0, 1 ml aliquots of bacterial suspension were transferred into 1.5 ml microcentrifuge tubes and aqueous solutions of peptides - magainin 2, buforin II, pleurocidin and D-LAK120-AP13 were added at the following concentrations: 250 μg/ml, 125 μg/ml, 62.5 μg/ml, 15.6 μg/ml, 3.9 μg/ml and incubated for 30 min at 37 °C . In order to be able to monitor the microbial recovery and growth, 10 μl of each suspension was sampled in 190 μl fresh medium onto a 96-well microplate. The OD_{620} was measured at time 0 and after 4 h of incubation at 37 °C . The microcentrifuge tubes were centrifuged at 5000 × g for 5 min and the bacterial pellets were snap frozen in liquid nitrogen, lyophilised and kept at -20 °C until further use. Pellets from triplicate tubes were combined for subsequent HR-MAS analysis. Each challenge was independently repeated nine times.

5.2.3 HR-MAS NMR

High-resolution magic angle spinning (HR-MAS) experiments were performed on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm $^{1}H/^{13}C$ HR-MAS probe. The lyophilised cell pellets were thawed at room temperature, transferred to an NMR rotor inserts and rehydrated with 30 μl of D_2O 2 hours before the acquisition. 1D spectra were recorded at a constant temperature of 310 K with magic angle spinning applied at 5 kHz. 1D ^{1}H spectra were recorded using a standard cpmgpr1d spin echo pulse (cpmgpr; Bruker) with water presaturation during recycle delay of 1 second and a total of 128 scans were acquired.

The spectral width was 16.02 ppm and ^{1}H 90 pulse length was 7.81 μsec. The free induction decay was multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. Phase correction was performed manually and automatic baseline correction was applied. A total of 120 samples were analysed with between 6 and 13 samples per treated condition and 17 control samples (no AMP treatment). A number of 2D experiments were run to facilitate identification of the compounds: homonuclear J-resolved 2D correlation with presaturation during relaxation delay using gradients (J-Res; jresgpprqf), ^{1}H/^{13}C correlation via direct inept transfer, phase sensitive using states, with decoupling during acquisition (HSQC 13C; AA-hsqcwg-13C), 2D homonuclear shift correlation with presaturation during relaxation delay (COSY; cosyprqf) all acquired using standard Bruker pulse sequences. Spectra were Fourier transformed, manually phase and automatically baseline corrected and calibrated with 2,2,3,3-D4-3-(Trimethylsilyl) propionic acid sodium salt (TMSP-2,2,3,3-D4) with the reference signal at 0 ppm.

5.2.4 Assignment

Resonances were assigned based on J-coupling partners revealed by COSY, multiplicities derived from JRES, statistical correlation spectroscopy (STOCSY) [94] and both ^{1}H and ^{13}C chemical shifts with reference to the *E. coli* metabolome database [58].

5.2.5 Multivariate data analysis

Spectra were analysed by PCA and OPLS-DA. Details of the method can be found in Chapter 2.3 on page 35 and the software manual can be found in Appendix A on page 177. First, the spectra were aligned to the reference peak and spectral regions such as water and reference peak (4.8 ppm and 0 ppm, respectively) and regions of no interest and/or no spectral information were removed. Spectra were then normalised using probabilistic quotient normalisation (PQN) [124] and autoscaled but not bucketed. Cross-validation was performed where 66 % of the samples were used as a training set and the remaining 33 % as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class, and that at least one sample from each class was present in the test set. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1-score used to choose the number of components, with the additional

constraint to use a maximum of 10 components. This double cross-validation was repeated 2000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This leads to 3 $\times$ 2000 models, each of these models leads to a point on the scores plot, but loadings and weights are presented as averages over all these models. The chosen number of components minus one was then used as an OPLS filter and a PLS-DA analysis with two components was carried out on the filtered data to yield one predictive and one orthogonal component. Finally, this procedure was repeated with randomly generated class assignments to provide a reference value for Q^2. In each case, genuine or permutated class assignments, the Q^2 value quoted is the mean of all models. Back-scaled loadings plots [113] were used to identify resonances with high variance and high weight, therefore the discriminating resonances, and verified against the peak intensity of the original spectra after PQN normalisation. Freely available MultiExperiment Viewer (MeV) which is a part of the TM4 Microarray Software Suite [133] was used for hierarchical cluster (HCL) analysis and generation of heatmaps. The euclidian distance algorithm was used to compute the differences between two gene expression levels (metabolite level changes) and the average linkage method was used to define the distances.

5.2.6 Scanning and transmission electron microscopy

Both SEM and TEM were used to examine the structural changes in bacteria induced by AMPs. Samples for the imaging were prepared in parallel with the samples used for HR-MAS NMR and hence represent bacteria in stationary phase. SEM sample preparation was performed by the author, whereas TEM samples were prepared by Centre for Ultrastructural Imaging (CUI) at King's College London. For SEM, the pellet obtained after centrifugation was fixed in 25 μl of 2.5 % (v/v) glutaraldehyde in 0.2 M sodium cacodylate buffer and kept at 4 °C until further use. In 24-well tissue culture plates 20 μl aliquots of vortexed bacterial pellet was smeared on 12 mm round poly-L-lysine (BD Biosciences, Bedford) cover slips with adjacent chambers filled with sufficient amount of 0.2 M sodium cacodylate to prevent drying of the slides and kept in a hydration chamber for 2 h. Cover slips were then washed with 0.2 M sodium cacodylate buffer followed by rinsing with 30 %, 70 %, 100 %, 100 %, and 100 % ethanol and incubating for 10 min between each wash. Hexamethyldisilazane (HMDS) was used for drying of the specimen by washing cover slips in 50/50 100 % ethanol/HMDS for 10 min followed by the final wash in HMDS for 10 min. The coverslips with dehydrated cells were mounted on the specimen stubs and

sputter coated with gold. Micrographs were acquired with FEI Quanta 200F FEG scanning electron microscope. Bacterial pellets for TEM processing were prepared as described above. Cells were pelleted by centrifugation and the pellet was post fixed in 1 % osmium tetroxide in 0.1 M phosphate buffer for 60 min an RT. The pellet was dehydrated by exposure to a graded series of ethanol (10 %, 70 % for 10 min each) followed by four washes in 100 % ethanol for 15 min each. Next, the pellet was subjected to two washes in propylene oxide, 10 min each. Tubes containing pellets were constantly rotated during the washes and the following procedures and the washes were performed in the fume hood. The supernatant was removed and the pellet placed into a mixture of 50 % resin and propylene oxide for 90 min and transferred to 100 % resin overnight before polymerisation at 60 °C for 24 hours. The resin blocks were sectioned with Leica Ultra-cut ultramicrotome to semi-thick sections (0.75 μM- 2 μM) and stained with toluidine blue and used to determine the areas for thin sectioning (90 nm). The sections were then placed onto 150 mesh copper grids coated with pioloform support film. Grids were then stained with uranyl acetate and lead citrate before viewing on Hitachi H7600 transmission electron microscope. For both techniques, around 15 images were taken for each treatment. The following magnifications were used and images were selected that are representative of the effect observed: 700 ×, 5000 ×, 12000 ×, 25000 ×, 70000 ×.

5.2.7 GeneChips

GeneChip experiments were performed using the Affymetrix® (Santa Clara, CA) *E. coli* Genome 2.0 Array with effective, response inducing, sub-MIC AMP concentrations determined from the HR-MAS metabolomic study; pleurocidin 62.5 μg/ml, buforin II 250 μg/ml, magainin 2 125 μg/ml and D-LAK120-AP13 15.6 μg/ml. Each array includes approximately 10,000 probe sets for all 20,366 genes present in four strains of *E. coli* over the entire open reading frame (ORF); K12 (MG1655 laboratory strain), CFT073 (uropathogenic), 0157:H7-EDL953 (enteropathogenic) and O157:H7-Sakai (enteropathogenic). RNA was extracted using RiboPure and enriched using MICROBExpress Bacterial mRNA Enrichment Kit after the DNA digestion step (Life Technologies, Paisley, UK). At each step the quality of RNA was assessed using Pico100 (Picodrop Ltd, Hinxton, UK). cDNA was synthesised from mRNA and purified using Qiagen MinElute PCR (Qiagen, Manchester, UK). cDNA was then fragmented and labeled using terminal transferase and biotinylated Affymerix® GeneChip labelling reagent according to the manufacturer's instructions. Fragmentation and labeling were

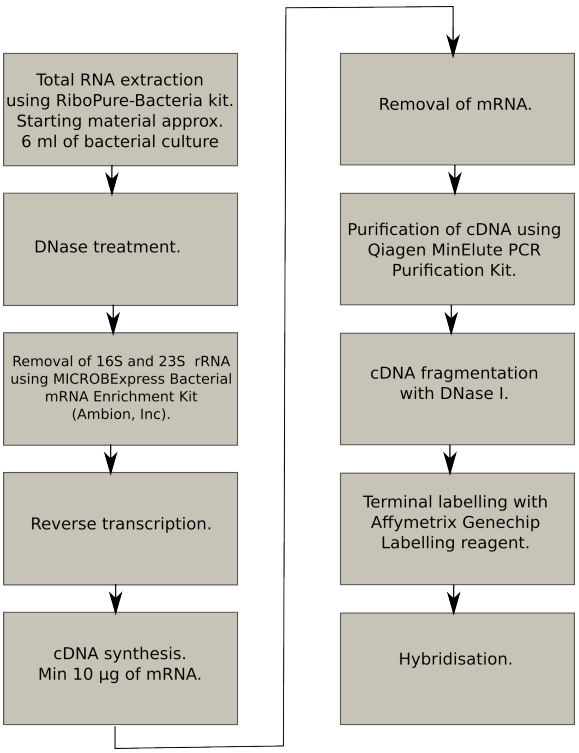


Figure 5.3: ***Flow diagram of steps in target preparation for Affymetrix® GeneChip system.***

assessed with the 2100 Bioanalyzer (Agilent Technologies, Wokingham, UK) to obtain the size distribution and yield. cDNA was kept at -80 °C until microarray hybridisation. Target hybridisation, scanning and Affymetrix® data analysis and formatting was performed by the Genomic Centre at King's College London. Hybridisation of the target to the GeneChip was prepared according the standard Prokaryotic Target Hybridisation protocol according to the manufacturer's instructions. The efficiency of the hybridisation step was assessed by examining hybridisation of Poly-A controls provided for the Affymetrix® GeneChip. Arrays were scanned on an Affymetrix® GCS3000 microarray system and image acquisition, quantification and data analysis were performed using Affymetrix® Command and Expression Console Software. Data were normalised using the Robust Multi-array Average (RMA) algorithm built into Expression Console. Preselection of gene lists for each treatment was performed using Qlucore Omics Explorer (Qlucore AB, Lund, Sweden). First, ANOVA across all samples identified the twenty most differentially expressed genes according to each replicated treatment. These were then assessed by principal component analysis to confirm that independently replicated experiments produced consistent results. Signal intensities for gene expression were then averaged across technical duplicates/triplicates and log transformed. For the gene annotation enrichment analysis, differentially

expressed genes in treatment versus control samples were selected by a paired, homoscedastic t-test with a significance cutoff of $p < 0.05$ and lists for the four AMP treatments were then compared using Venny [321]. Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-1703. To better understand the differences between the effects of the four treatments, significance thresholds that identified the approximate top 200 - 250 differentially expressed genes were selected; $p \leq 0.0184$ for buforin II and D-LAK120-AP13, $p \leq 0.0425$ for pleurocidin and $p \leq 0.078$ for magainin 2. These lists were analysed using the GOEAST Gene Ontology Enrichment Analysis Software Toolkit where the Benjamini-Hochberg option was selected allowing an FDR up to 15 % [322]. Discriminating metabolite changes, identified from HR-MAS NMR, were then mapped onto the KEGG pathway using BioCyc Omics Data Analysis (http://ecocyc.org/PToolsWebsiteHowto.shtml#omicsDataAnalysis) and genes related to given metabolic pathway checked against consistently differentially expressed genes, whether or not they had passed the significance test described above.

5.2.8 Multiparameter viability assays

In order to assess the functionality and cellular integrity of bacteria the following viability assays have been used: membrane potential assay, esterase activity assay and BacLight Live-Dead stain for microscopy [323].

As previously, *E. coli* NCTC 9001 were grown from glycerol stocks in Muller-Hinton broth overnight at 37 °C without shaking until an OD_{620} of 1.0 was reached. 1 ml aliquots of culture were challenged for 30 min with four peptides at and below the threshold concentrations established with NMR. Cells were then harvested by centrifugation at 5,000 × g for 5 min and washed in 50 mM phosphate buffer (pH 7.0). For BacLight Live/Dead stain cells were diluted to 4×10^8 CFU/ml, whereas for the remaining assays cells were diluted to 2×10^8 CFU/ml. All experiments were performed at room temperature. Negative controls were obtained either by treatment with 70 % isopropanol for 10 min and removed by centrifugation at 5,000 × *g* for 5 min and re-suspension in PBS, or by heat killing at 85 °C for 10 min on a heat block. Assays were performed in black, flat bottom, 96-well plates and read on a Synergy HT multi-mode microplate reader (BioTek, Winooski, VT).

Membrane potential

25 mg of dye $DiBAC_4$ (Anaspec, Fremont, CA) was reconstituted in 2.42 ml ethanol to obtain a 20 mM stock solution which was stored at -20 °C . The stock was diluted further with water to working concentration of 12.5 μM immediately before use. 20 μlof 12.5 μM dye was added to a 96-well plate, covered by 180 μlbacterial suspension in PBS and mixed. The plate was incubated in the dark for 5 minutes and fluorescence emission was measured (excitation 485 nm, emission 535 nm). Since membrane damage leads to higher fluorescence intensity, values were background corrected and expressed as a reciprocal before being normalised with untreated cells defined as being 100 % and isopropanol treated cells defined as 0 %.

Esterase activity

5 mg of esterase substrate 5,6-carboxyfluorescein diacetate (CFDA) was dissolved in 1.086 ml dimethyl sulfoxide (DMSO) to obtain 10 mM stock kept at -20 °C . Stock was diluted 40 × in water immediately before use to obtain working concentration of 250 μM, which was pre-aliquoted to a 96-well plate. 180 μlof bacterial suspension in PBS was added to the plate and mixed with the detection solution. The plate was incubated in the dark for 30 minutes with occasional shaking and fluorescence emission measured (excitation 485 nm, emission 535 nm).

LIVE/DEAD BacLight

This assay was performed by the author and the images were taken and processed by Dr Garrit Koller. LIVE/DEAD BacLight kit (Life Technologies, Paisley, UK) was used to measure membrane integrity. Harvested cells were reconstituted with saline and 3 μlof the dye mixture (1.5 μlof SYTO9 (3.34 mM) and 1.5 μlof propidium iodine (20 mM)) was added to each 1 ml of bacterial suspension and mixed. Tubes were incubated for 15 minutes in the dark with occasional shaking and fixed with 20 % paraformaldehyde (PFA) and kept at 4 °C . Specimens were viewed on an Olympus BX60 microscope fitted with an Andor Ultrahigh-resolution CCD setup. A × 20 oil immersion lens was used to obtain a 200 μm field width. Excitation and emission filters were 480/520 nm and 515/560 nm respectively.

5.2.9 MIC testing

Parent strain BW25113 and Keio knockout strains [324] for Δ*yejF*, Δ*yjjB*, Δ*yohN*, Δ*yrdB*, Δ*metB*, Δ*cyoA*, Δ*cyoC*, Δ*cyoD*, Δ*speB*, and Δ*argR* were obtained from the Coli Genetic Stock Center (Yale University, New Haven, CT). The activities of the peptides were assessed in planktonic suspension in polypropylene 96 well plates (Greiner Bio-one, Frickhausen, Germany) according to a modified broth dilution assay [325]. Bacteria were grown without shaking in 50 ml Mueller-Hinton (MH) broth at 37 °C . Peptides (pleurocidin, magainin 2 and LL-37) were tested in duplicates with two rows allocated for each peptide. In each of columns 2 - 11, 50 μlof MH broth was added under sterile conditions. In the first column of each row, 50 μlof 256 μg/ml stock peptide solutions, prepared in distilled water, were added and then the broth from the second column was pipetted into the first column and thoroughly mixed before being deposited again in the second column. This process was repeated throughout the tray providing a twofold dilution of peptide with each row. Bacteria with an OD_{620} of 0.001 were then added to each well in volumes of 50 μlgiving a further twofold dilution and a final volume of 100 μlper well. The final column was used either as sterility control (100 μlbroth) or negative control (no peptide). Plates were incubated overnight at 37 °C and the OD_{620} read. Growth curves prepared from duplicates were fitted to determine the peptide concentration required to inhibit growth by 50 % (MIC_{50}). The MIC_{50} quoted for each peptide (Table 5.1 on page 120) is an average value from at least two independent repeats.

5.3 Results

5.3.1 ^{1}H HR-MAS NMR metabolomics reveals threshold AMP concentration

^{1}H HR-MAS NMR metabolomics allowed successful differentiation between bacterial responses to a challenge with four distinct AMPs (Table 5.1 on page 120). All of the cationic amphipathic AMPs chosen for this study were of similar length and were all C-terminally amidated with nominal charge ranging from +4 to +9. Unlike in conventional microdilution assays [325] that were used in MIC testing (MIC_{50} for all peptides given in Table 5.1 on page 120), for the AMP challenge experiments higher bacterial cell densities (8×10^8 CFU/ml) were required at the stationary phase. At the higher bacterial titre, the relative potency of AMPs was similar, however, the effect of the four AMPs was somewhat different from

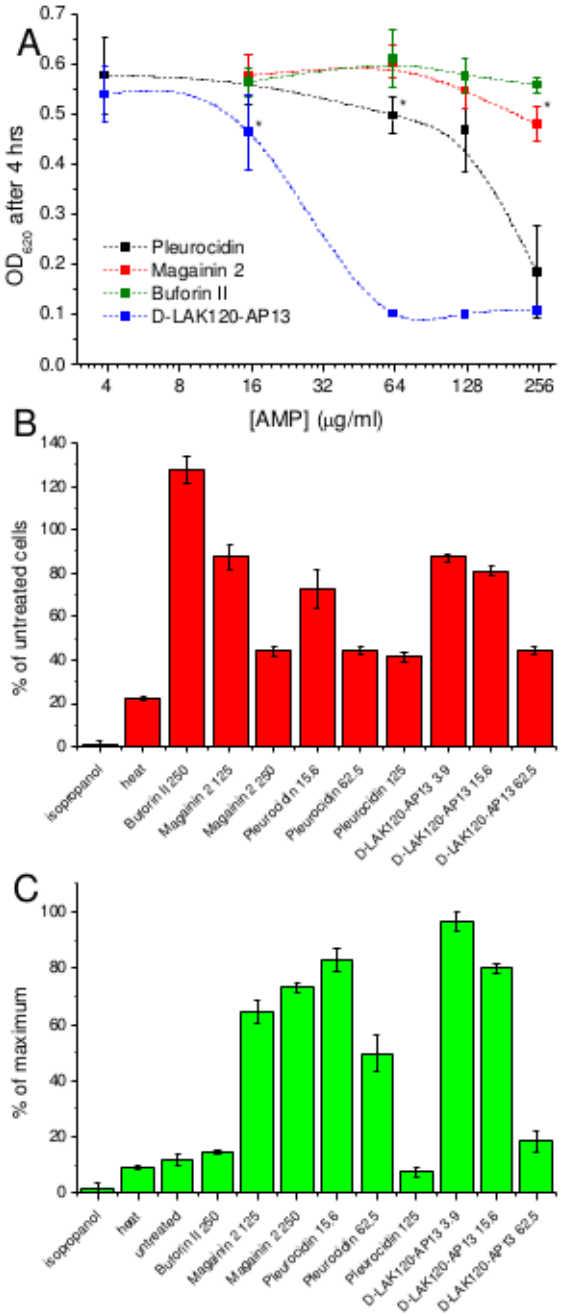


Figure 5.4: ***AMP challenge of*** **E. coli** ***NCTC 9001:*** *overnight cultures were challenged with increasing amounts of each of four AMPs for 30 minutes and the recovery of aliquots added to fresh media was measured after 4 hours incubation at 37°C (A). ∗ indicates the peptide concentration causing a significant ($p < 0.1$) reduction in OD_{620} relative to the lowest peptide concentration used. Membrane potential of challenged bacteria as measured by the voltage sensitive dye $DiBAC_4$ and expressed as a percentage of untreated cell (B). Esterase activity determined by cleavage of 5,6-carboxyfluorescein diacetate expressed as a percentage of the maximum observed activity (C). Peptide concentrations are given in µg/ml.*

the minimum inhibitory concentrations (MICs), with D-LAK120-AP13 having a substantially greater effect on bacterial numbers as detected in the challenge and recovery assay (Figure 5.4 on the preceding page (A)). Magainin 2 and Buforin II had insufficient inhibitory effect for a MIC to be determined at the higher titre. Nevertheless, the amount of peptide causing a significant reduction in bacterial re-growth was established with D-LAK120-AP13 effective at 15.6 μg/ml, pleurocidin at 62.5 μg/ml and magainin 2 at 125 μg/ml required for a significant effect. Buforin II had no observable effect on bacterial re-growth at any of the peptide concentrations tested (Figure 5.4 on the previous page (A)). Additionally, a multi-parameter assay was taken to assess the effect of peptide challenge on membrane potential (Figure 5.4 on the preceding page (B)), esterase activity (Figure 5.4 on the previous page (C)) and membrane integrity in the challenged stationary phase bacteria. The results suggested that only the concentrations of pleurocidin and D-LAK120-AP13 above the threshold were lethal. A dose dependent response to each of the four AMPs was observed, however even at higher peptide concentrations, the membrane potential was not completely lost while the esterase activity was higher than in the untreated cells, which is a typical response in *E. coli* to exposure to sublethal stress [323].

Both 1D and 2D NMR spectra acquired for the metabolomics study were of good quality and major metabolites were assigned as seen from assigned COSY 5.5 on the following page and HSQC 5.6 on page 133.

One dimensional ^{1}H NMR spectra were obtained for all samples. PCA identified outlier spectra resulting from either poor baseline or signal to noise, whereas either PLS regression analysis or a series of OPLS-DA tests were used to interrogate the spectra. The lowest concentration of each AMP that caused a significant change in the spectra relative to spectra from untreated cells was determined using PLS-DA in a step-wise manner. The resulting Q^2 values are presented in Table 5.2 on page 138. An arbitrary value ≥ 0.6 was taken to select a reliable model where the AMP challenge has a significant effect. This value was also compared with the reference Q^2 that was obtained with randomly assigned class identifiers in order to show a model with no effect. 2D scores plots that resulted from each of the cross validated OPLS-DA analyses are shown in Appendix B on page 185 (Figure B.9 on page 195, B.10 on page 196, B.11 on page 196). Scores plots for the threshold concentrations of AMP and for the highest AMP concen-

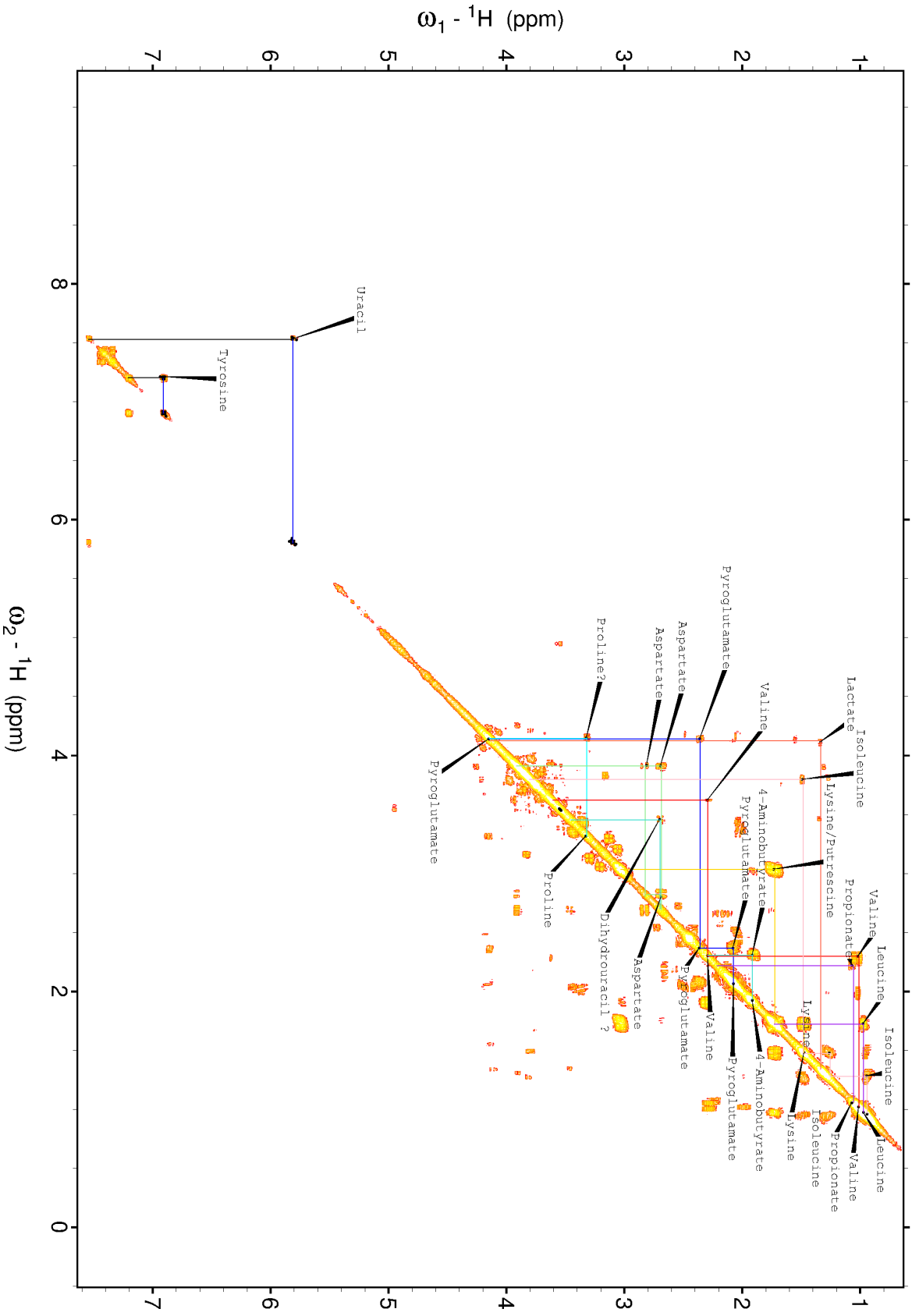


Figure 5.5: ***Representative 2D COSY spectrum of*** **E. coli** ***NCTC 9001*** *challenged with pleurocidin at the threshold concentration and the assigned metabolic compounds.*

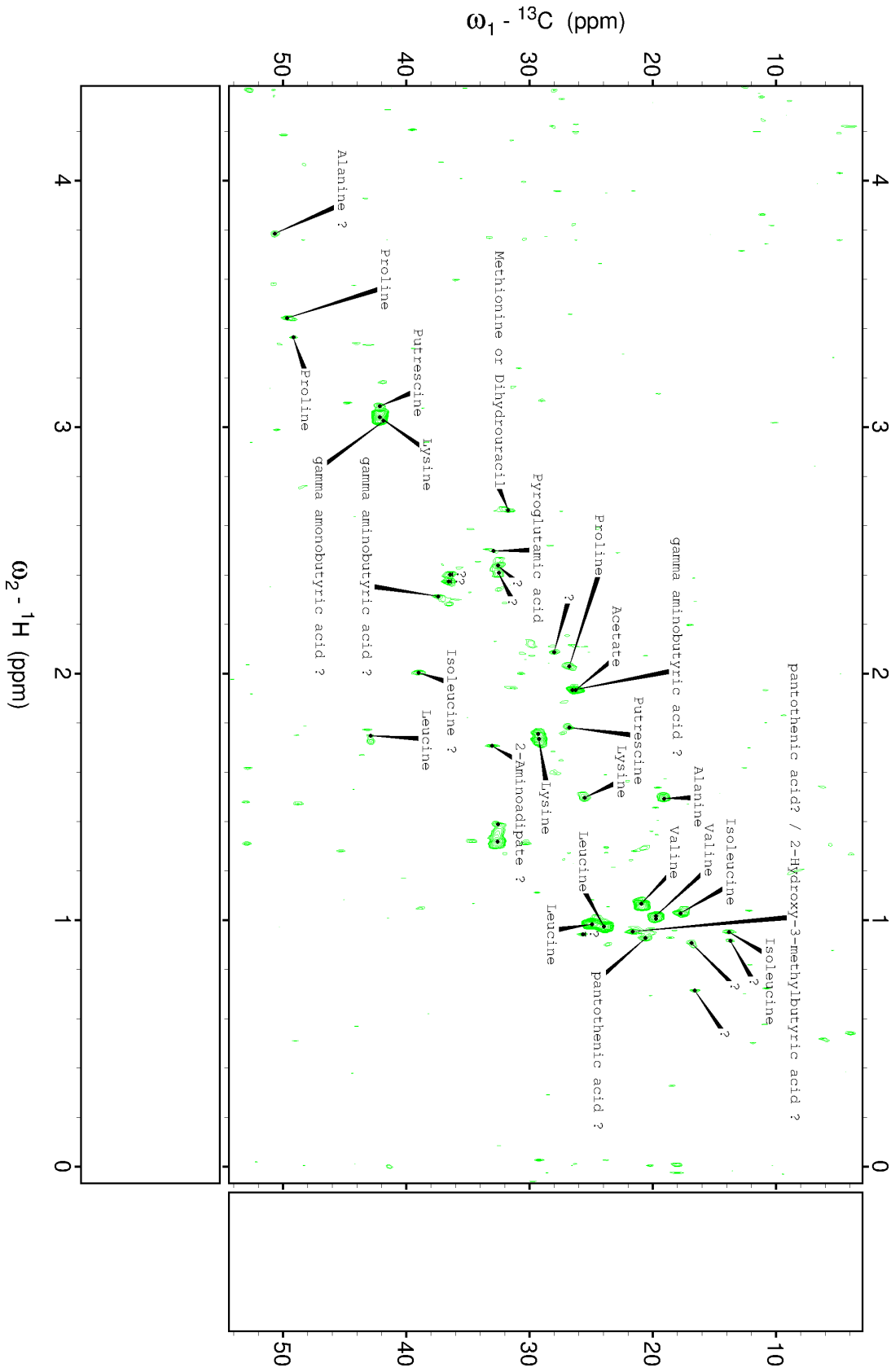


Figure 5.6: ***Representative 2D*** ^{1}***H*** **-** ^{13}***C HSQC spectrum of*** **E. coli** ***NCTC 9001*** *challenged with pleurocidin at the threshold concentration and the assigned metabolic compounds.*

tration of buforin II are presented in Figure 5.11 on page 140. A clear separation of the OPLS-DA scores that was obtained at the indicated AMP concentration and was accepted as an indication of a threshold concentration of AMP needed for a response detectable in the bacterial metabolomes. This threshold concentration varied considerably for the four AMPs and was directly related to the apparent antibacterial efficacy as detected by MIC_{50}. The growth curves (Figure 5.4 on page 130 (A)) indicated that bacterial growth remained greater than 50 % of maximum after challenge with AMPs at the threshold concentration, while esterase activity was increased relative to untreated cells (Figure 5.4 on page 130 (C)) and membrane potential was not completely lost (Figure 5.4 on page 130 (B)). Taken together the threshold value corresponded to a sub-lethal AMP concentration, therefore the NMR metabolomic technique identified conditions where metabolically active *E. coli* were responding to the AMP challenge without simply reporting on bacterial cell death. Comparing the backscaled loadings for each binary comparison between untreated bacteria and those challenged with each AMP using OPLS-DA, identified metabolites whose differing intensities correlated with the effect of each AMP. A hierarchical cluster analysis was used to reveal variation in metabolite levels when comparing bacteria challenged at threshold concentrations (Figure 5.7 on the next page) and comparing above and below threshold concentrations (Figure 5.8 on page 136). Both common and AMP specific changes in *E. coli* metabolite levels were observed in response to challenge with the four AMPs. Notably, the hierarchical analysis grouped the peptides according to their potency as seen in both dendrograms. Assigned metabolites were used as an input to a network pathway analysis, which was conducted using MetaboAnalyst [101, 326]. The magnitude and direction of changes in metabolite levels were not taken into account. Metabolic pathways were identified according to p-values obtained from pathway enrichment analysis and pathway impact from pathway topology (Figure 5.22 on page 153). Changes in alanine, aspartate and glutamate metabolism had the greatest impact and were a common feature of challenge with all four peptides with changes in pyruvate, butanoate and arginine/proline metabolism highlighted according to the distinct challenges.

The dynamic response of *E. coli* NCTC9001 to challenge with pleurocidin or magainin 2 was assessed over a period of 2 hours at the following intervals: 5 minutes, 15 minutes, 60 minutes and 120 minutes. The OPLS-DA scores plots (Figure 5.9 on page 137) indicate that a response to AMP challenge at the level of

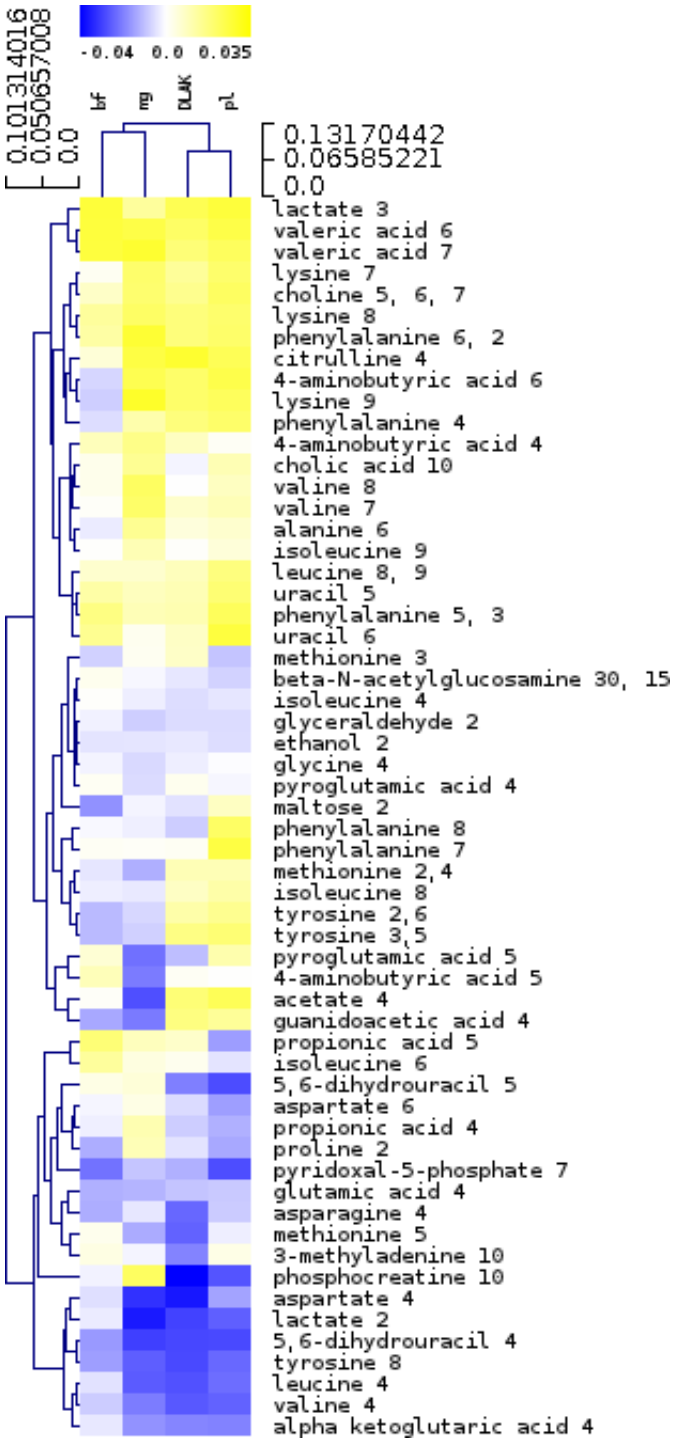


Figure 5.7: ***Hierarchical clustered heatmap*** *comparing loadings obtained from cross-validated OPLS-DA comparing untreated bacteria with AMP at the threshold concentrations indicated above.*

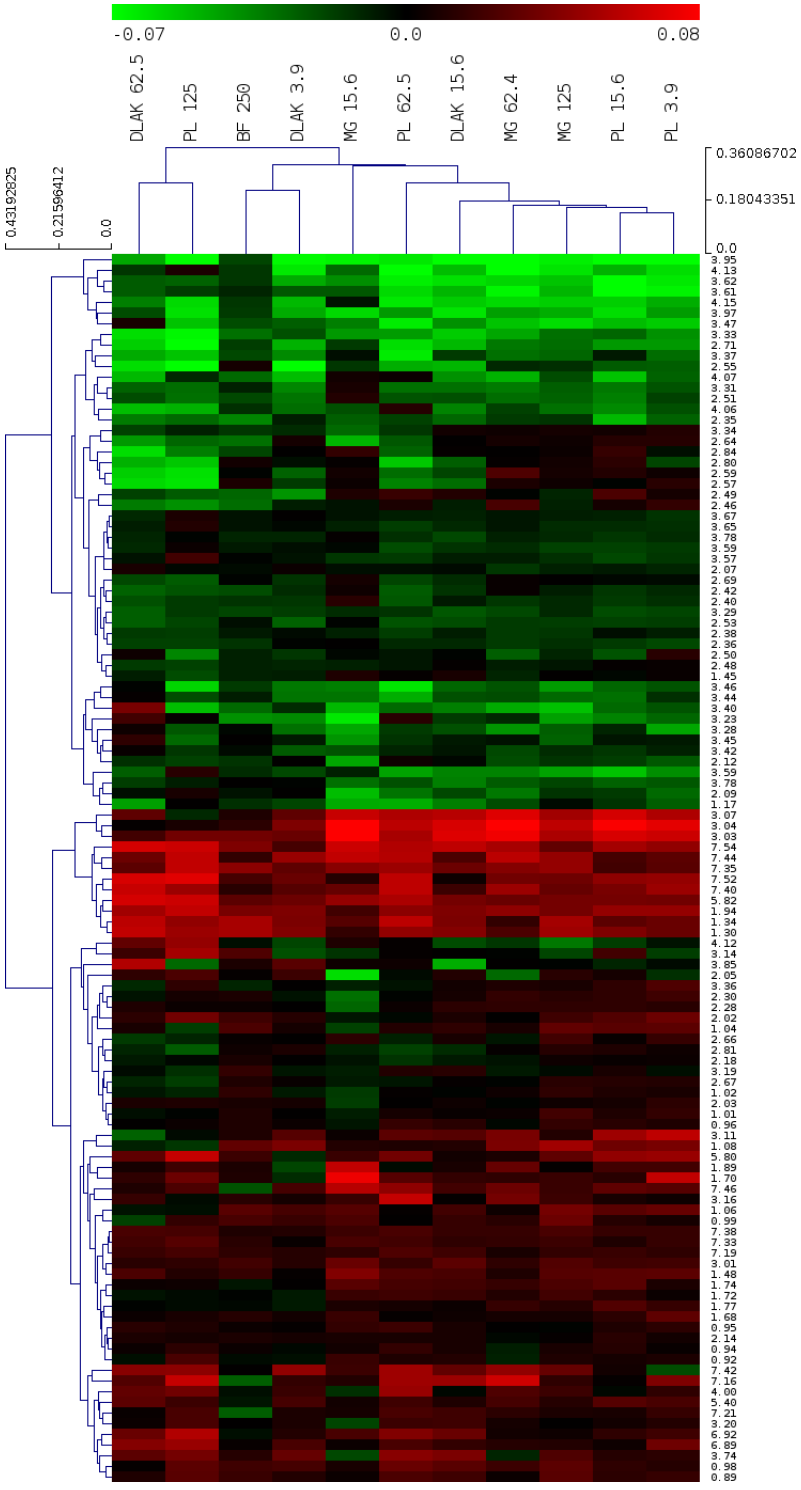


Figure 5.8: ***Hierarchical clustered heatmap*** *comparing loadings obtained from cross-validated OPLS-DA comparing untreated bacteria with AMP at the above and below threshold concentrations.*

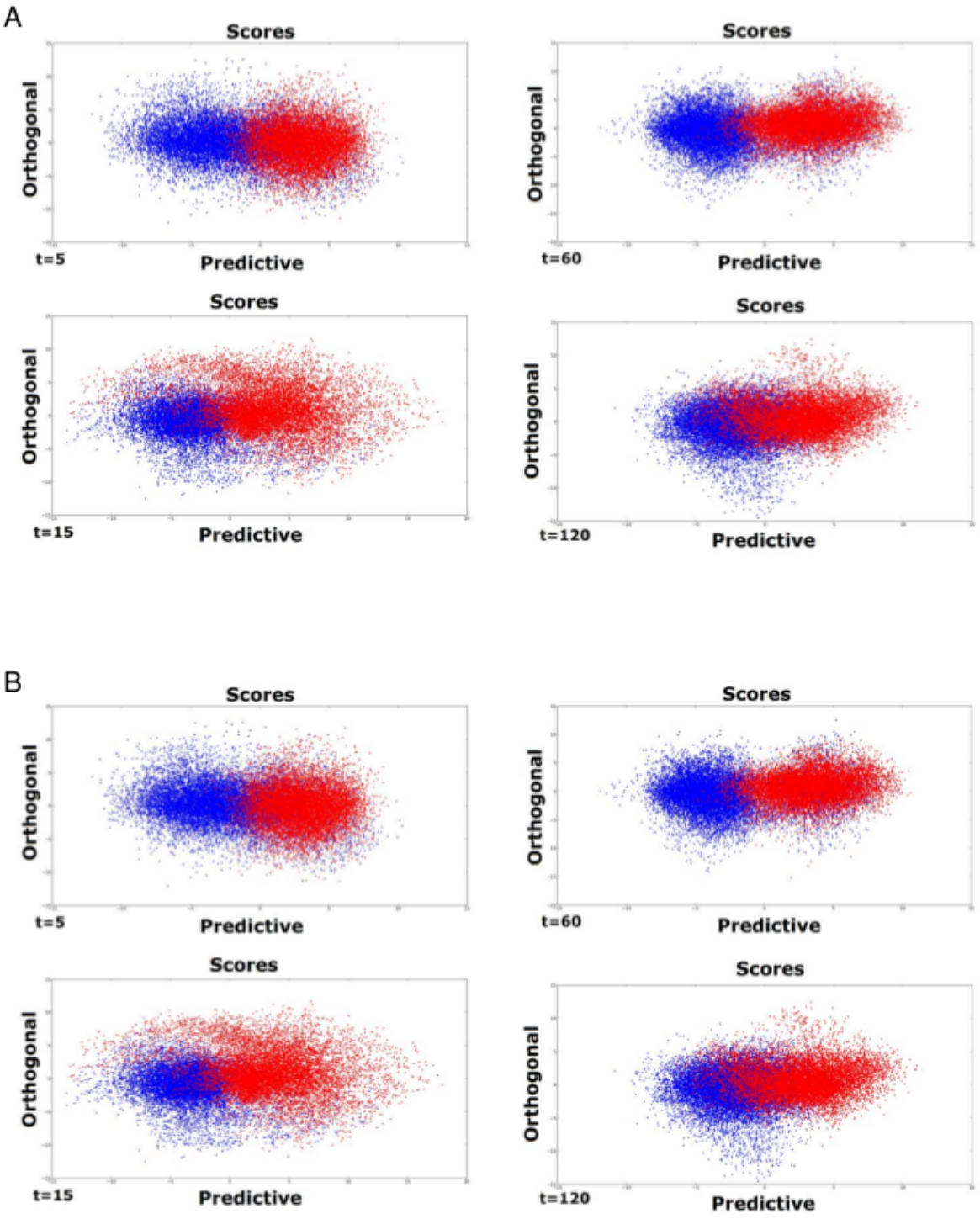


Figure 5.9: ***Comparison of OPLS-DA scores plot from 2000 cross-validated models for bacteria treated with 125 μg/ml magainin 2 (A) or 62.5 μg/ml pleurocidin (B), against untreated control at t = 5, t = 15, t = 60, and t = 120 minutes.***

Table 5.2: ***Predictive Q^2 values for OPLS-DA models.*** *Q^2 values for cross validation performed with permutated classes are provided in parentheses.*

[AMP] (μg/ml)	Q^2			
	Pleurocidin	Magainin 2	Buforin II	D-LAK120-AP13
3.9	0.32 (-0.29)	n.d	n.d.	0.37 (-0.30)
15.6	0.53 (-0.31)	0.29 (-0.36)	n.d.	0.59 (-0.28)
62.5	0.81 (-0.41)	0.20 (-0.31)	n.d.	0.81 (-0.31)
125	0.80 (-0.29)	0.68 (-0.34)	n.d.	0.83 (-0.26)
250	n.d.	n.d.	-0.30 (-0.39)	n.d.

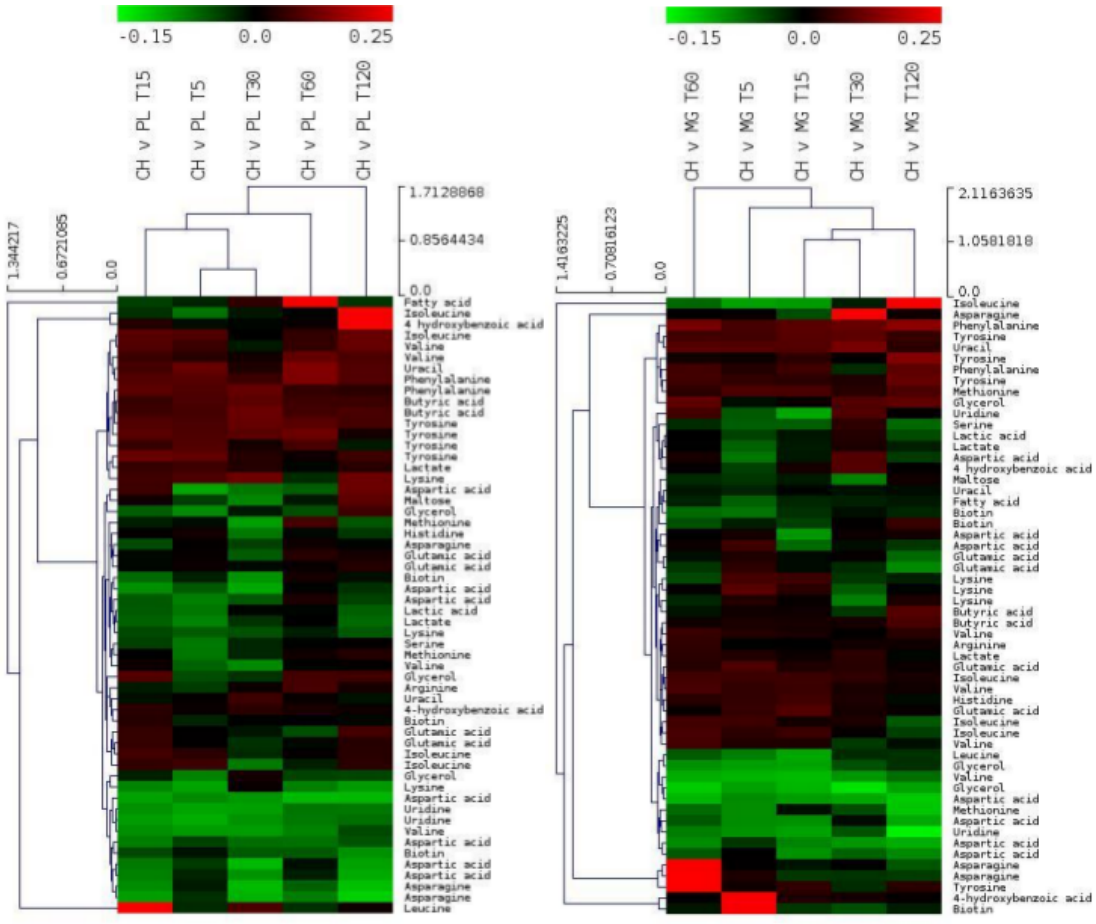
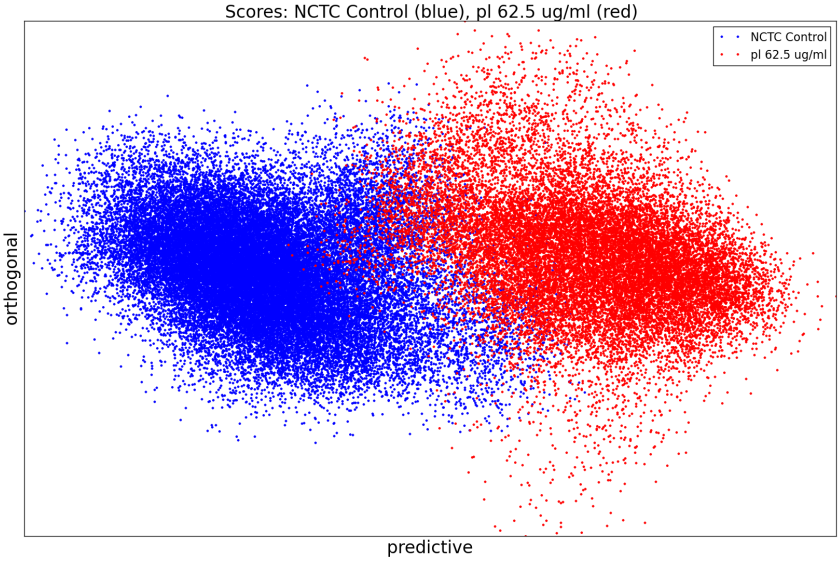
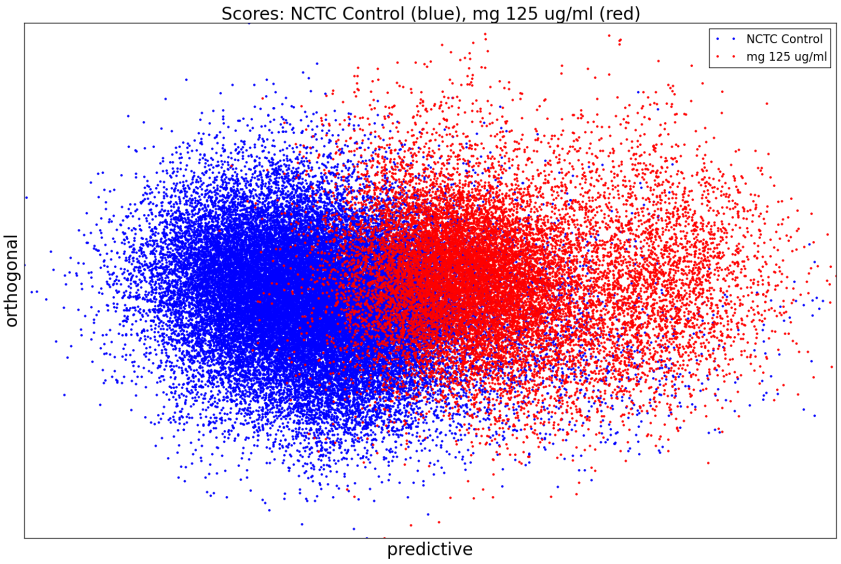


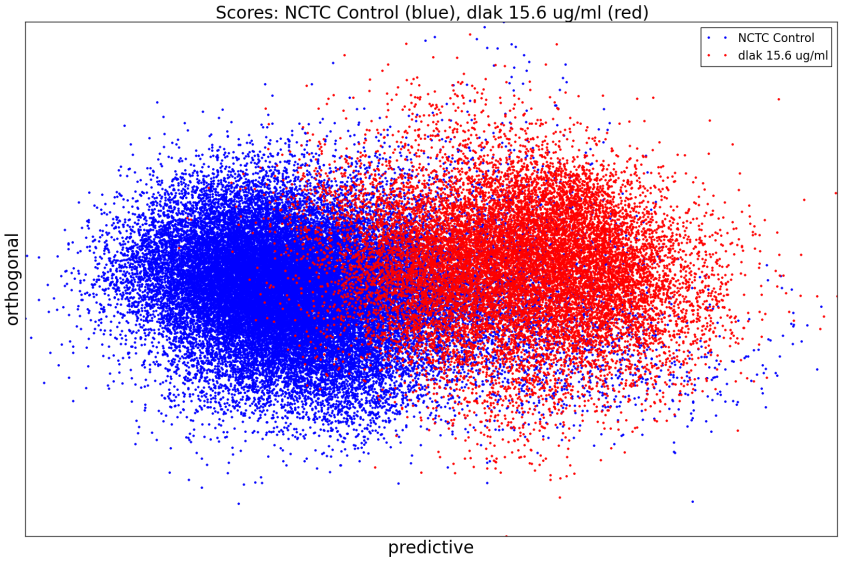
Figure 5.10: ***Hierarchical cluster analyses of metabolic responses to pleurocidin (left) and magainin 2 (right) challenge recorded for five different incubation periods.*** *The responses are broadly similar over time but, in particular for pleurocidin, there is a suggestion that a second phase can be detected after 30 minutes.*


Scores: NCTC Control (blue), pl 62.5 ug/ml (red)
NCTC Control
pl 62.5 ug/ml
orthogonal
predictive

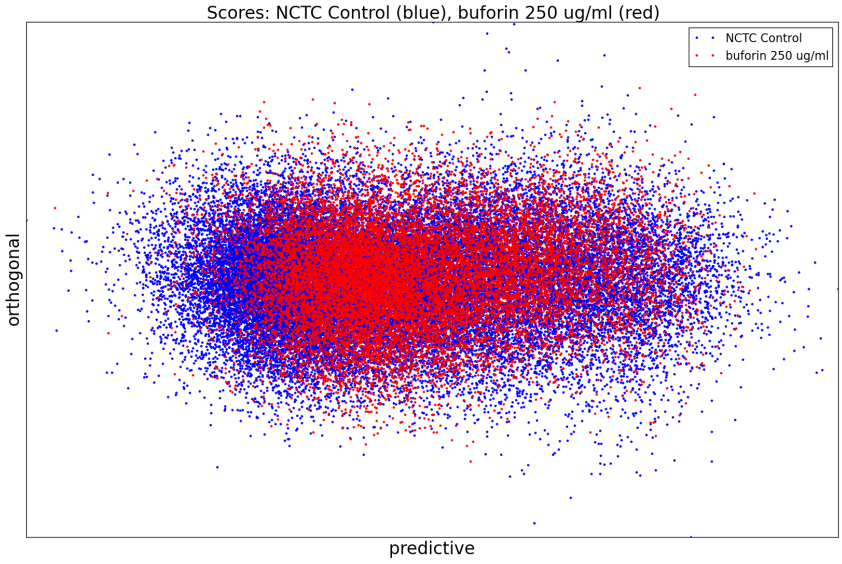
A


Scores: NCTC Control (blue), mg 125 ug/ml (red)
NCTC Control
mg 125 ug/ml
orthogonal
predictive

B



C



D

Figure 5.11: ***Metabolomic analysis by*** ^{1}***H HR-MAS NMR of lyophilised, stationary phase*** **E. coli** ***cell pellets.*** *OPLS-DA scores plots are shown for challenge at the following threshold concentrations; pleurocidin at 62.5 µg/ml (A), magainin 2 at 125 µg/ml (B), D-LAK120-AP13 at 15.6 µg/ml (C) and buforin II at 250 µg/ml (D).*

the metabolome can be detected throughout the period tested. However, when the backscaled loadings were compared in a hierarchical cluster analysis (Figure 5.10 on page 138), modest but notable differences in the affected metabolites were discerned. This suggested that the bacterial response detected beyond an hour after challenge is characteristically distinct from that probed within the first 30 minutes. These conditions - 30 minutes incubation at the determined threshold concentration - were therefore used for subsequent electron microscopic and transcript profiling analyses of samples prepared in parallel to those used above.

5.3.2 Scanning and transmission electron microscopy identifies differences in the response to each AMP

Changes in *E. coli* internal or external morphology in response to challenge with AMP were monitored respectively using transmission and scanning electron microscopy (TEM/SEM) at either one or four times the sub-inhibitory AMP threshold concentration known to induce a metabolomic response (Figure 5.14 on page 144 to 5.19 on page 149 and B.1 on page 187 to B.8 on page 194). The bacterial response to each AMP challenge varied considerably and was in qualitative agreement with the metabolomic study; buforin II had no noticeable effect when compared with untreated bacterial cell controls (Figure 5.16 on page 146 and 5.17 on page 147 D/E), with each of the three other AMPs inducing substantial changes to external and/or internal morphologies. For magainin 2, a regular, almost circular nucleoid condensation was observed in some, but not all, cells (Figure 5.3.2 A, B.7 on page 193) while some impairment of cell division was evident with extended rods observed (Figure 5.15 on page 145 A). Pleurocidin also induced nucleoid condensation but this was much more widespread; observed throughout the bacterial cell population (Figure B.4 on page 190/ B.5 on page 191). This was accompanied by some possible protein aggregation and the production of large amounts of a fibrous material (Figure 5.14 on page 144 B). In addition to the production of the fibrous material, SEM identified moderate vesicle production, a known envelope stress response in Gram-negative bacteria [327]. Finally, D-LAK120-AP13 induced dramatic changes in both the internal (Figure 5.14 on page 144 C) and external *E. coli* morphologies (Figure 5.16 on page 146 C). Extensive release of outer membrane vesicles was evident which was coincident with a loss of the normal rod shape, consistent with bacteria budding prematurely (Figure 5.16 on page 146 C). Inside bacterial cells, extensive nucleoid condensation and protein aggregation was observed throughout the bacterial cell

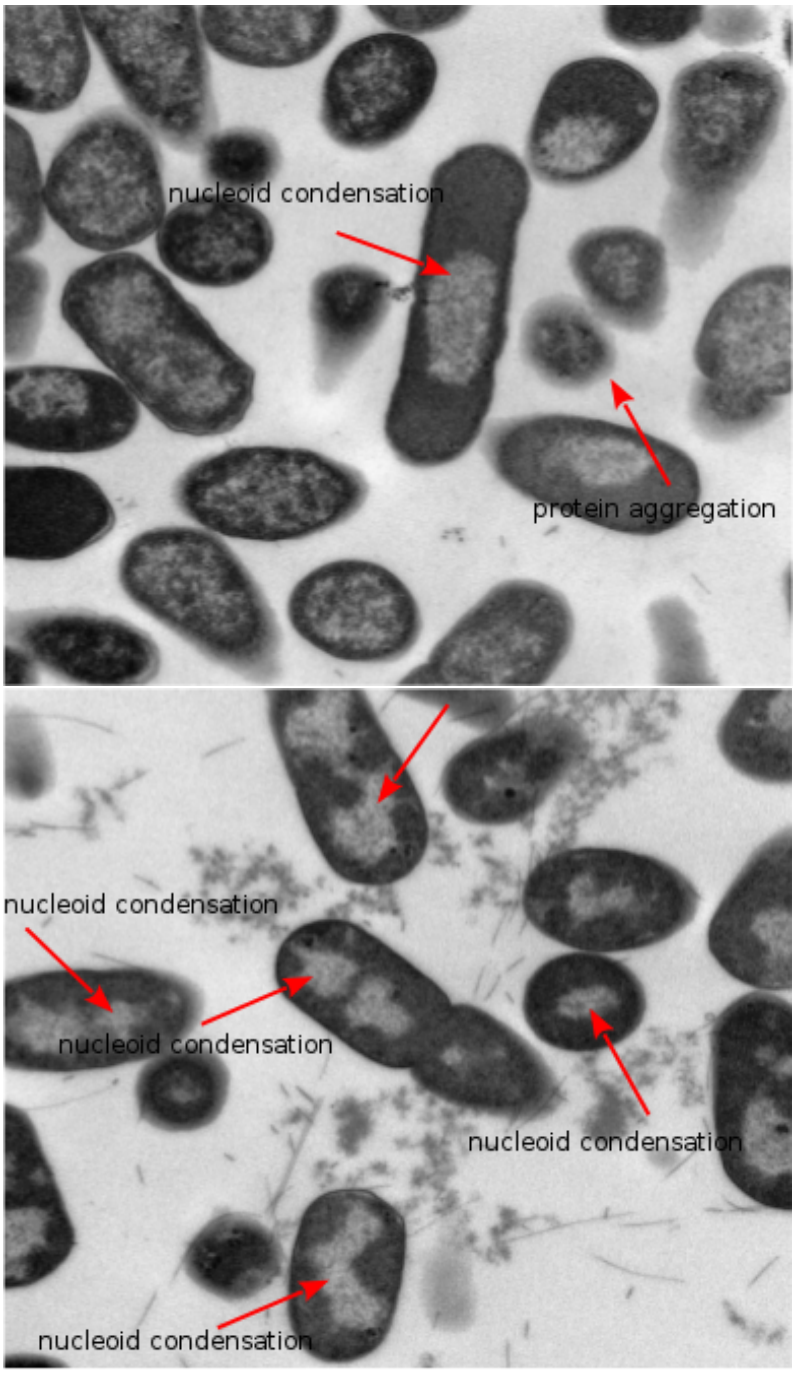
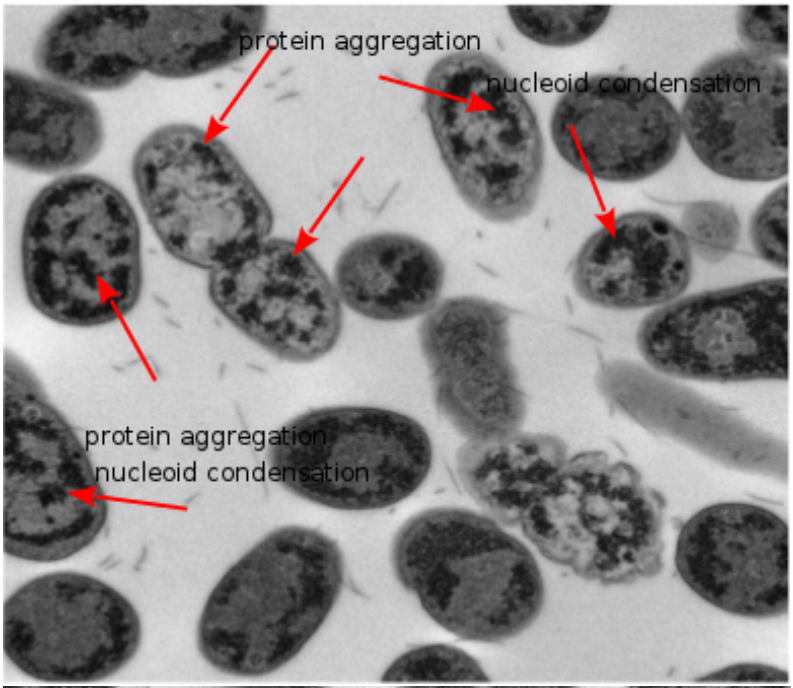


Figure 5.12: ***Transmission electron micrographs at × 25,000 magnification of AMP challenged* E.coli *NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomics study; 250 µg/ml magainin 2 (A), 125 µg/ml pleurocidin (B).*

C



D

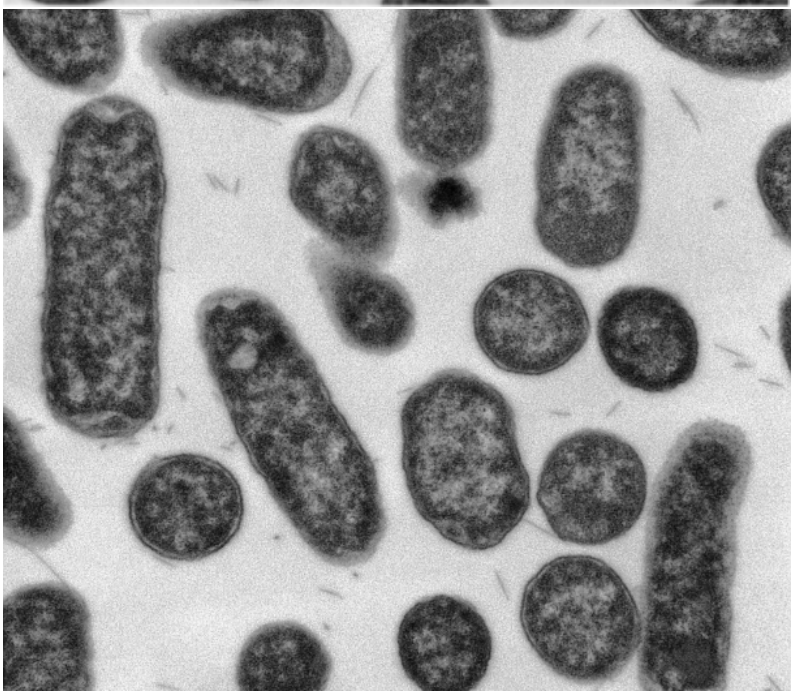


Figure 5.13: ***Transmission electron micrographs at × 25,000 magnification of AMP challenged*** **E.coli** ***NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomics study; 62.5 µg/ml D-LAK120-AP13 (C), 250 µg/ml buforin II (D).*

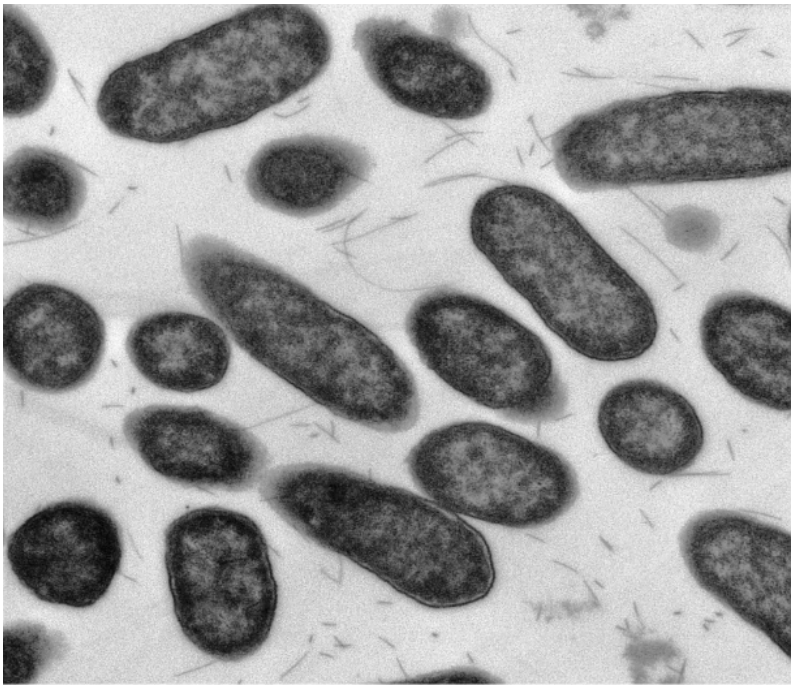
E 

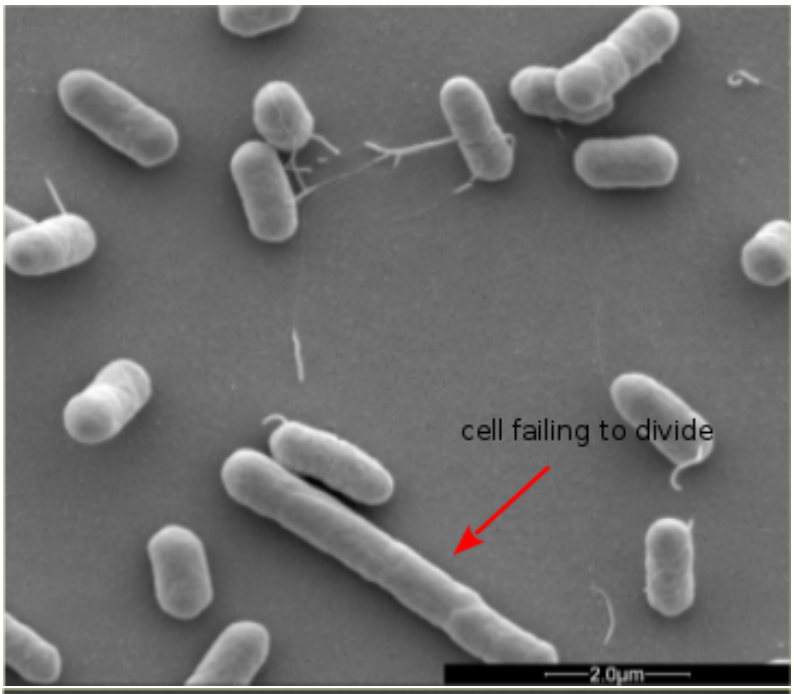
Figure 5.14: ***Transmission electron micrographs at × 25,000 magnification of untreated*** **E. coli** ***NCTC 9001.***

population (Figure 5.14 C; Figure B.2 on page 188, B.3 on page 189). Taken together, although there were some qualitative similarities in the response of *E. coli* cells to each of the three more potent AMPs, markedly distinct responses to each peptide were observed overall. Transmission electron micrographs obtained at higher magnification and with AMP added at a concentration above the detected threshold value indicated that, for all four peptides, the bacterial envelope remained intact and no release of cell contents was apparent (Figure 5.19 on page 149).

5.3.3 Global transcriptome response

The response of *E. coli* to challenge with the four AMPs was then probed at the level of the transcriptome. Transcript profile changes in the NCTC 9001 strain, a clinical isolate from a patient with cystitis, were monitored using the *E. coli* Genome 2.0 Array where four strains including laboratory, uropathogenic and enteropathogenic strains are featured. Due to the high degree of similarity between strains, in the majority of cases, a single probe set represents the equivalent ortholog in all four strains. All genes that are subsequently described in detail are found in both laboratory (K12 substr. MG1655) and uropathogenic (CFT073) strains with the majority also found in the two enteropathogenic

A



B

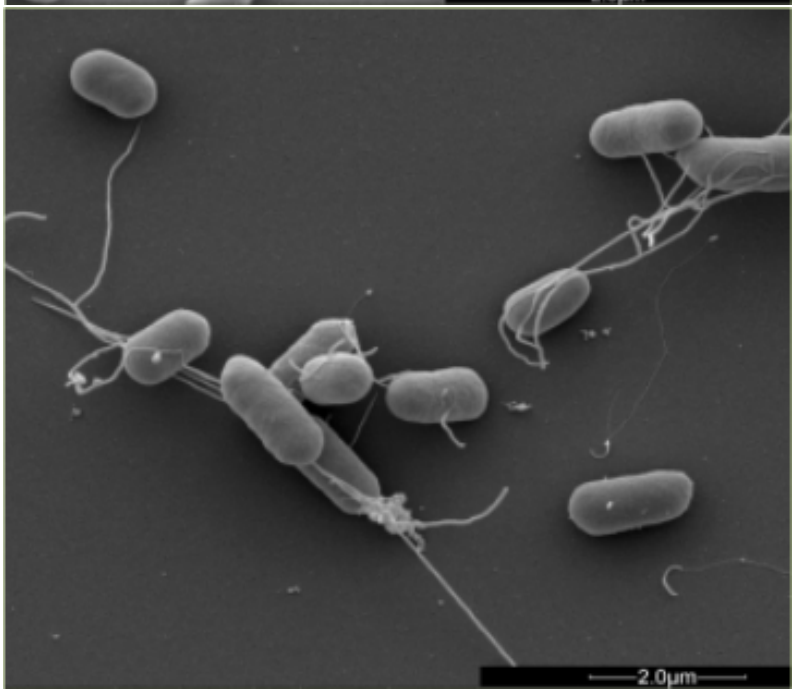
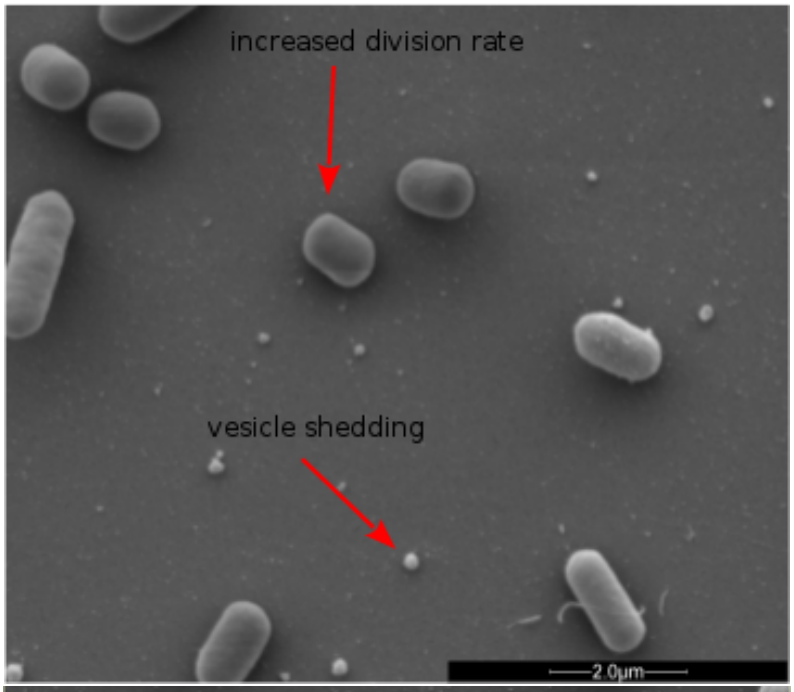


Figure 5.15: ***Scanning electron micrographs at × 25,000 magnification of AMP challenged* E. coli *NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 250 µg/ml magainin 2 (A), 125 µg/ml pleurocidin (B).*

C



D

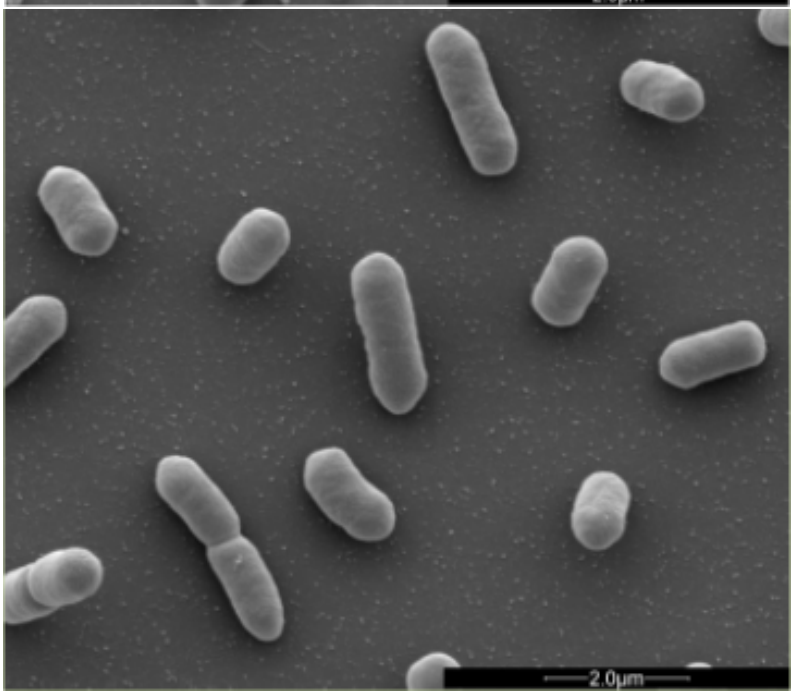


Figure 5.16: ***Scanning electron micrographs at × 25,000 magnification of AMP challenged* E. coli *NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 62.5 µg/ml D-LAK120-AP13 (C), untreated* E. coli *NCTC 9001 (D).*

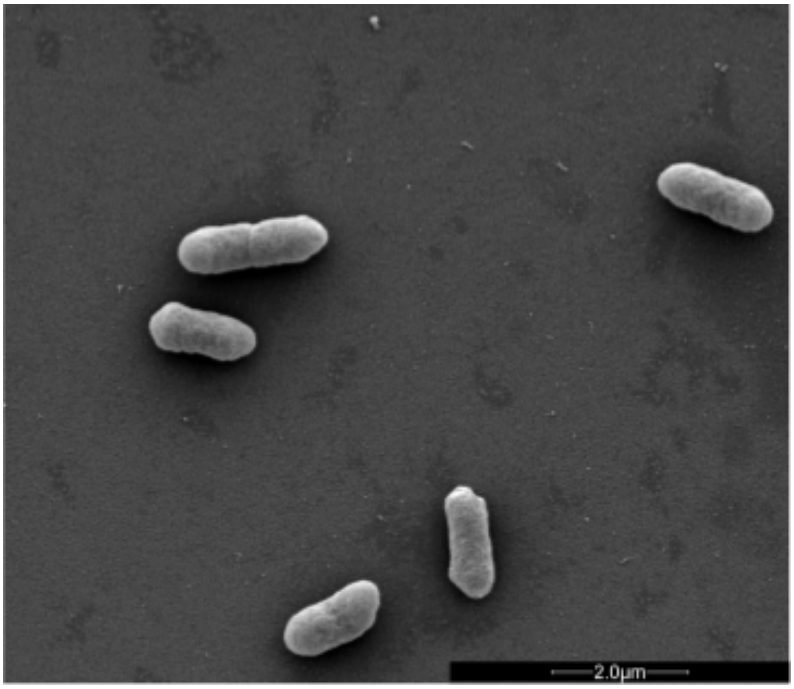
E 

Figure 5.17: ***Scanning electron micrographs at × 25,000 magnification of AMP challenged*** **E. coli** ***NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 250 µg/ml buforin II (E).*

strains. PCA of the twenty most differentially expressed genes across all groups showed the three independent replicates of each condition clustered together indicating the AMP challenge and transcript profiling assay were reproducible (Figure 5.20 on page 150). Further analysis, where either an arbitrary significance level ($p \leq 0.05$) for differential gene expression or manual manipulation of significance levels leading to an optimal separation by principal components, generated lists of differentially expressed genes related to each treatment. *E. coli* genomes commonly encode between approximately 4,200 and 5,500 protein coding genes [328, 329]. Of the approximately 10,000 probe positions, between 139 and 632 differentially expressed unique genes ($p \leq 0.05$) were detected for each treatment following challenge with AMP at the threshold concentration eliciting a bacterial response. This corresponds to 2.5 - 15.0 % of the available genome. Magainin 2 induced differential expression of only 139 genes which contrasted with the much greater number of genes whose expression was altered in response to challenge with either buforin II or D-LAK120-AP13; 625 and 632 respectively. Pleurocidin induced differential expression of 298 genes. The distribution of differentially expressed genes according to each AMP treatment is represented in a Venn diagram and reveals that the vast majority (76.3 %) are specific to each of the four

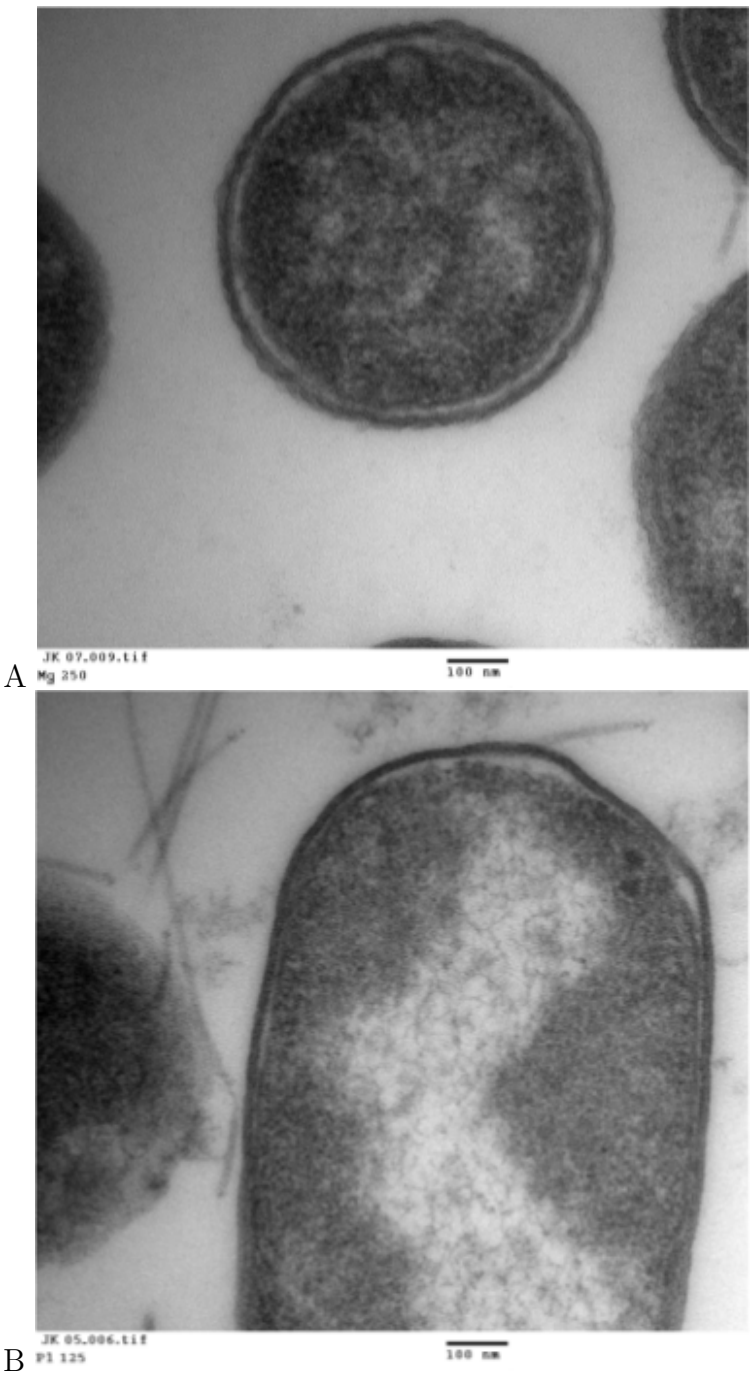


Figure 5.18: ***Transmission electron micrographs of AMP challenged*** **E. coli** ***NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 250 μg/ml buforin II (A),125 μg/ml pleurocidin (B).*

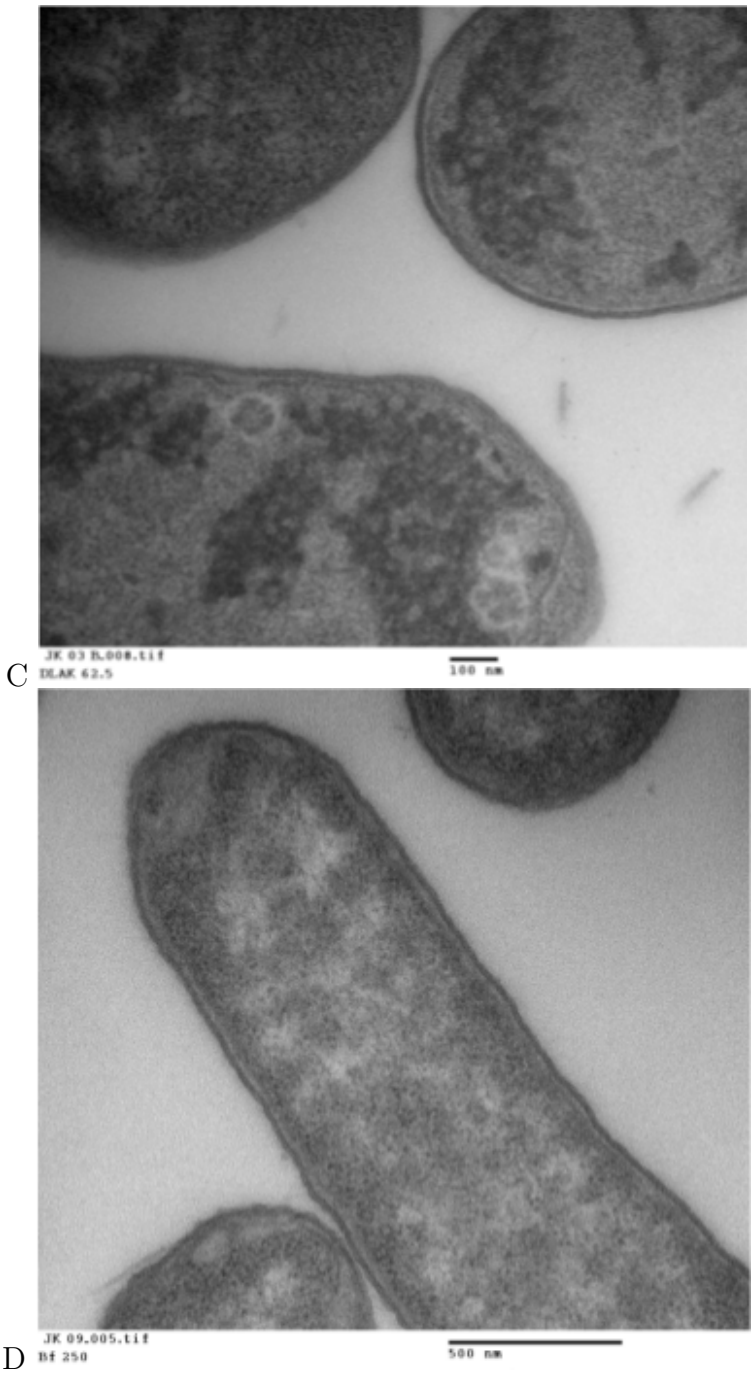


Figure 5.19: ***Transmission electron micrographs of AMP challenged*** **E. coli** ***NCTC 9001.*** *Bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 250 µg/ml magainin 2 (C) and 62.5 µg/ml D-LAK120-AP13 (D).*

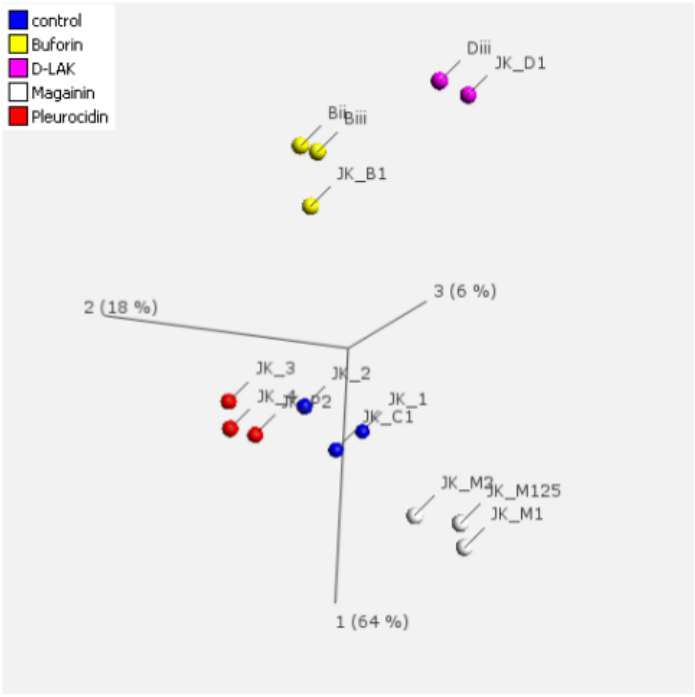


Figure 5.20: ***Output from Qlucore Omics Explorer showing three dimensional Principal Component Analysis of 20 most differentially expressed genes across all 14 GeneChips for*** **E. coli** ***NCTC 9001*** *as detected by the GeneChip* E. coli *Genome 2.0 Array. Bacteria were challenged for 30 minutes with AMPs at the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study; 250 µg/ml buforin II, 62.5 µg/ml pleurocidin (B), 125 µg/ml magainin 2 (C) and 15.6 µg/ml D-LAK120-AP13 (D). The axes (1, 2, 3) relate to principal component 1 (PC1) , PC2 and PC3 respectively and indicate how much variance is explained by each of these first three principal components. The plot indicates the reproducibility of the transcript profiling experiment by showing that variance in the 20 most different differentially expressed genes is closely related to the AMP challenge applied.*

AMP challenges (Figure 5.21 on the following page). Only 32 differentially expressed genes, 2.4 % of the total, were common to at least three treatments while there was only one, *yjjB*, which was common to all four treatments. Qualitatively therefore, transcriptomic data supported the electron microscopy findings as, while common responses can be identified, the dominant impression was of a largely specific response to each AMP challenge.

Discriminating metabolite changes with the most impact (Figure 5.22 on page 153) were mapped onto their respective Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. This allowed identification of differentially

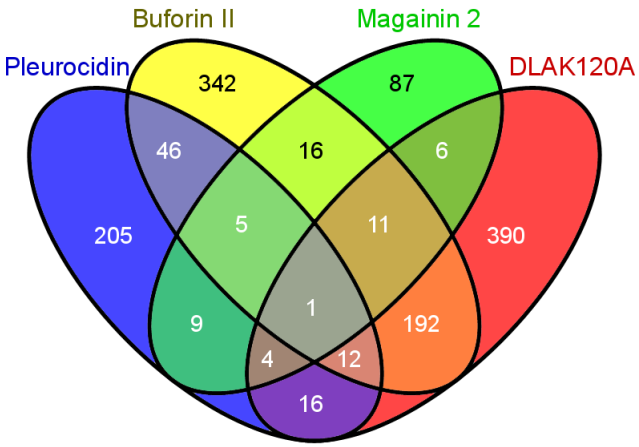
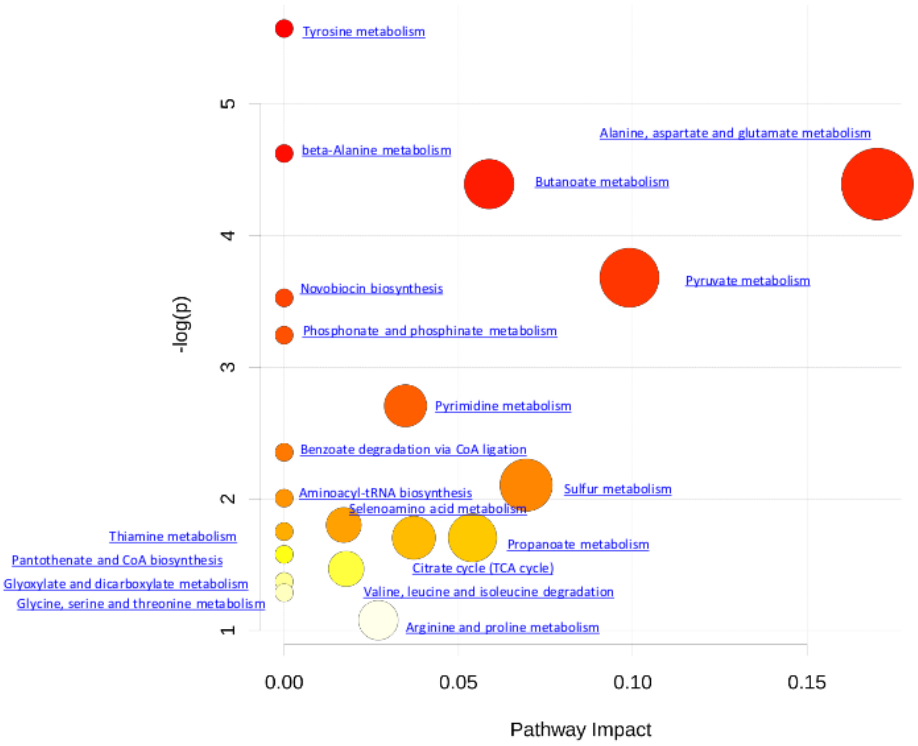
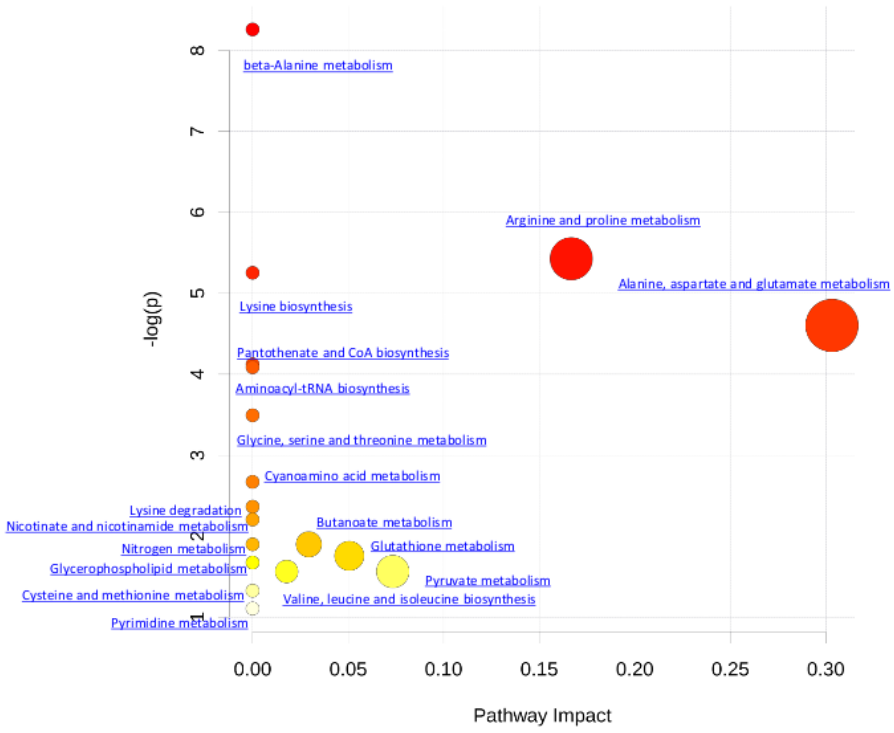


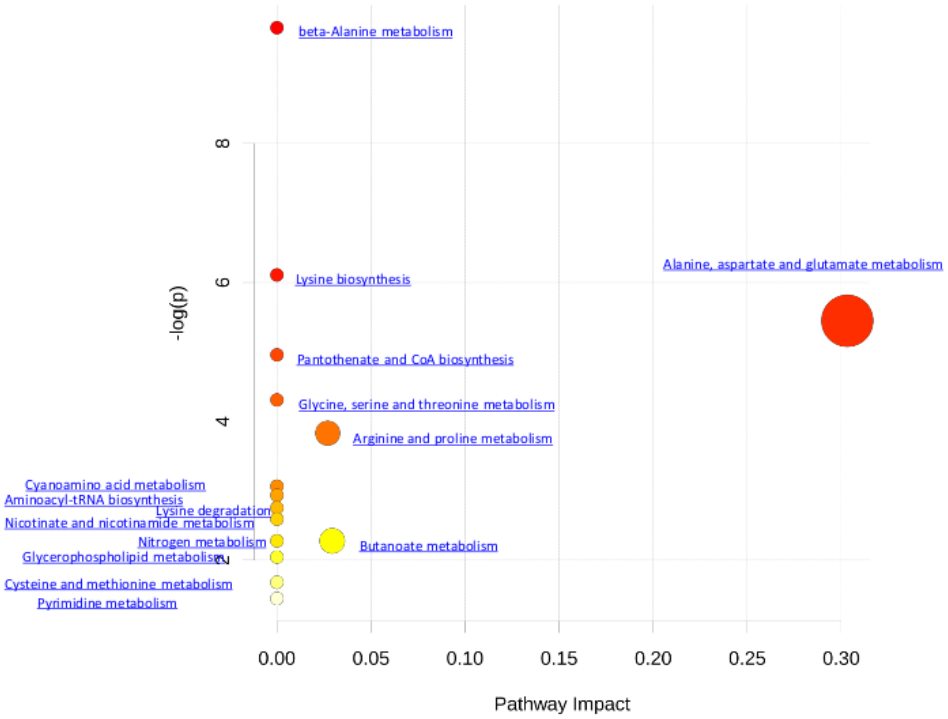
Figure 5.21: ***Transcript profiles and role of individual genes in response to AMP challenge.*** *Four way Venn diagram showing the distribution of differentially expressed genes detected by the GeneChip* E. coli *Genome 2.0 Array ($p \leq 0.05$) following challenge of stationary phase* E. coli *NCTC 9001 with each of four AMPs at subinhibitory concentrations known to elicit a bacterial response; pleurocidin at 62.5 μg/ml, magainin 2 at 125 μg/ml, D-LAK120-AP13 at 15.6 μg/ml and buforin II at 250 μg/ml. The entries in the Venn correspond to the number of affected genes.*


Tyrosine metabolism
beta-Alanine metabolism
Alanine, aspartate and glutamate metabolism
Butanoate metabolism
Pyruvate metabolism
Novobiocin biosynthesis
Phosphonate and phosphinate metabolism
Pyrimidine metabolism
Benzoate degradation via CoA ligation
Sulfur metabolism
Aminoacyl-tRNA biosynthesis
Selenoamino acid metabolism
Thiamine metabolism
Propanoate metabolism
Pantothenate and CoA biosynthesis
Citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
Valine, leucine and isoleucine degradation
Glycine, serine and threonine metabolism
Arginine and proline metabolism
-log(p)
1
2
3
4
5
0.00
0.05
0.10
0.15
Pathway Impact

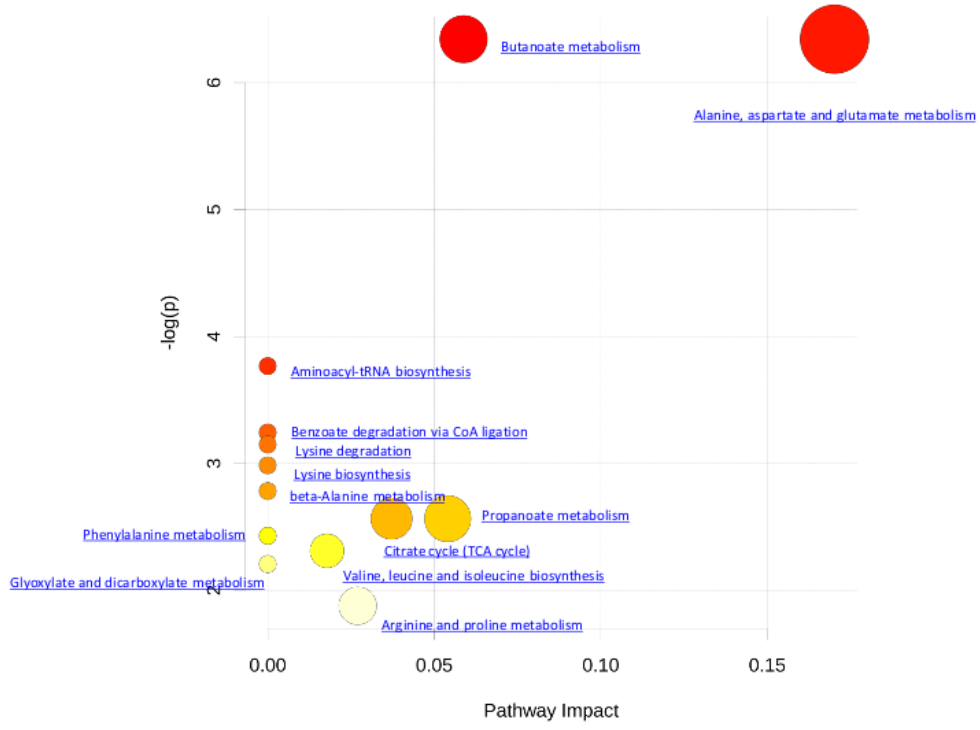
A


beta-Alanine metabolism
Arginine and proline metabolism
Alanine, aspartate and glutamate metabolism
Lysine biosynthesis
Pantothenate and CoA biosynthesis
Aminoacyl-tRNA biosynthesis
Glycine, serine and threonine metabolism
Cyanoamino acid metabolism
Lysine degradation
Nicotinate and nicotinamide metabolism
Butanoate metabolism
Nitrogen metabolism
Glutathione metabolism
Glycerophospholipid metabolism
Pyruvate metabolism
Cysteine and methionine metabolism
Valine, leucine and isoleucine biosynthesis
Pyrimidine metabolism
-log(p)
1
2
3
4
5
6
7
8
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Pathway Impact

B



C



D

Figure 5.22: ***Network pathway analysis by MetaboAnalyst software*** *showing matched pathways according to p-values from pathway enrichment analysis and pathway impact values from pathway topology analysis based on the identified NMR resonances distinguishing control from the treatment with 62.5 µg/ml pleurocidin (A), 125 µg/ml magainin 2 (B), 15.6 µg/ml D-LAK120-AP13 (C), 250 µg/ml Buforin II (D).*

expressed genes that potentially participate in mediating the response to AMP challenge. Changes in alanine, aspartate and glutamate metabolism were common to all four peptides and changes in expression of *gltX*, *dapA* and *metB*, coding for respectively glutamyl-tRNA synthetase, dihydropicolinate synthase and cystathionine gamma-synthase, were observed in the gene lists though these did not always satisfy the significance thresholds used above. Knockout mutants of *dapA* and *gltX* are not available from the Keio collection but Δ*metB* and five other knockout mutants (Δ*cyoA*, Δ*cyoC*, Δ*cyoD*, Δ*speB*, and Δ*argR* coding respectively for cytochrome o uniquinol oxidase subunits II, III and IV, agmatinase and arginine repressor), linked to changes in arginine/proline metabolism, were tested for altered sensitivity to AMP challenge though none was found.

Up-regulated in response to challenge by all four AMPs, *yjjB*, encodes a 157 amino acid, conserved, inner membrane protein predicted to have four transmembrane helices but with no known function. Of the five genes whose expression was generically affected by the three AMPs of natural origin, three were up-regulated in response to AMP challenge; *manA* codes for mannose-6-phosphate isomerise, *cysE* codes for a serine acetyltransferase and *yohN* codes for a 112 amino acid integral membrane protein annotated and established as a periplasmic modulator of nickel and cobalt efflux and renamed *rcnB* [329]. In contrast, *yejF*, part of an ABC transporter identified as a possible nickel, and probable microcin C transporter [330], and *yrdB*, which codes for a highly anionic, glutamine rich, 85 amino acid hypothetical protein from the DUF1488 superfamily, are down-regulated. Comparison of the growth of parent strain BW25113 and four knockout mutants (Δ*yejF*, Δ*yjjB*, Δ*yohN* and Δ*yrdB*) obtained from the Keio collection [324] confirmed *yohN* confers sensitivity to Co^{2+} and possibly Ni^{2+} (Figure 5.23 on the following page). The growth of these strains was also tested in the presence of AMPs (Figure 5.24 on page 157). While the MIC for pleurocidin was not affected by the presence of any of the four deletions, a modest but significant ($p < 0.05$) increase in sensitivity was observed for all four deletion strains when challenged by magainin 2. When the experiment was repeated with LL-37, an AMP of human origin, three of the deletions rendered the bacteria more sensitive while deletion of *yrdB* had no effect. Table 5.3 on page 156 lists genes related to metal binding that were differentially expressed, such as genes involved in nickel ion binding *nikA* and *nikB*, components of nickel ABC transporter, which were upregulated when challenged with pleurocidin and magainin 2, but downregulated in response to the D-peptide and buforin II. A member of two-component

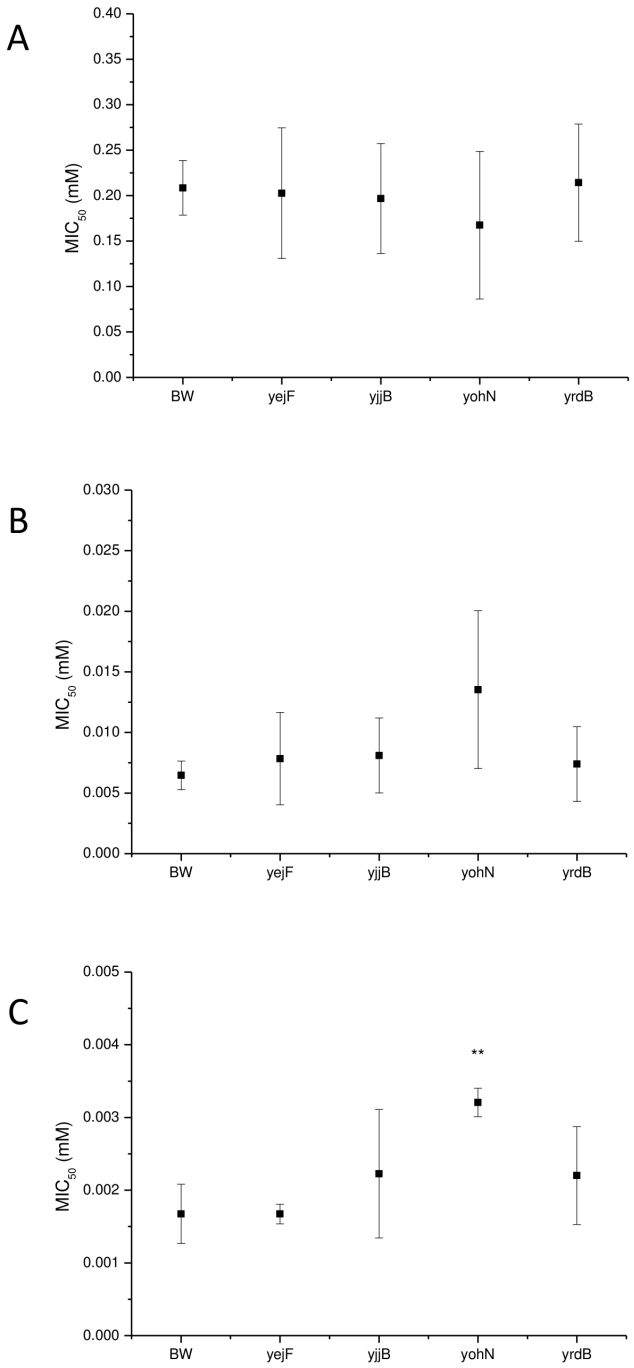


Figure 5.23: ***Sensitivity of Wild type and four mutants from the Keio collection to different cations:*** *(A) $MgCl_2$, (B) $NiCl_2$, (C) $CoCl_2$. (**) $p \leq 0.05$ relative to BW.* ΔyohN *confers sensitivity to Co^{2+} and possibly Ni^{2+}.*

regulatory system which mediates copper tolerance, *cusR* was upregulated across all challenges. Other genes involved in metal binding, which are considered iron uptake virulence factors found were: *fepC*, encoding iron-enterobactin transporter ATP-binding protein and *iucD* of the aerobactin operon which were substantially downregulated in response to all four peptides, *fepG* encoding iron-enterobactin transporter permease, which was upregulated for D-peptide and magainin 2, but downregulated for pleurocidin and buforin II. Also, both genes encoding highly conserved proteins IscU and IscA that are members of an operon iscSUA, which provides scaffold protein for assembly and transfer of iron-sulfur clusters were strongly repressed across all challenges. This is also evident from ontology search of biological processes (Figure B.12 on page 197) where iron-sulfur cluster assembly and metalo-sulfur cluster assembly can be found for bacteria challenged with pleurocidin.

Table 5.3: ***Differentially expressed genes related to metals and metal binding up- or downregulated in response to challenge with AMPs at sub-inhibitory concentration.*** *In bold* $p \leq 0.05$.

		Fold change			
Gene symbol	**Gene title**	**Pleurocidin**	**Buforin**	**DLAK**	**Magainin**
nikA	nickel-binding, heme-binding periplasmic protein	**2.45**	0.51	0.56	1.34
nikB	nickel transporter permease	**0.51**	0.78	0.61	**1.89**
iscA	iron-sulfur cluster assembly protein	**0.52**	0.40	0.56	0.93
iscU	iron-sulfur cluster assembly scaffold protein	0.54	0.41	0.57	0.86
fepC	iron-enterobactin transporter ATP-binding protein	1.08	1.06	1.19	**0.45**
fepE	ferric enterobactin transport protein	**0.21**	0.18	0.26	0.51
fepG	iron-enterobactin transporter permease	**0.84**	0.95	1.35	1.21
feoA	Ferrous iron transport protein	0.78	0.43	0.43	**0.32**
fdoH	Formate dehydrogenase-O iron-sulfur subunit	**0.31**	0.63	0.57	**0.45**
iucD	IucD protein	**0.50**	0.53	0.63	0.79

The ontological profile related to each challenge offers another view of how closely related the response to each AMP is to each other. Here, instead of comparing individual genes on the basis of their identity, the comparison is based on the cellular component, biological process or molecular function and is less affected by redundancy or more subtle changes in response and consequently better reflects the fundamentals of the bacterial response. Ontological analysis, which employed a Benjamini-Hochberg method to control false discovery rate (FDR) and displays statistically overrepresented, differentially expressed genes in a graphical format according to their relationships in a hierarchical tree, was carried out on gene lists comprising the 200 - 250 most differentially expressed

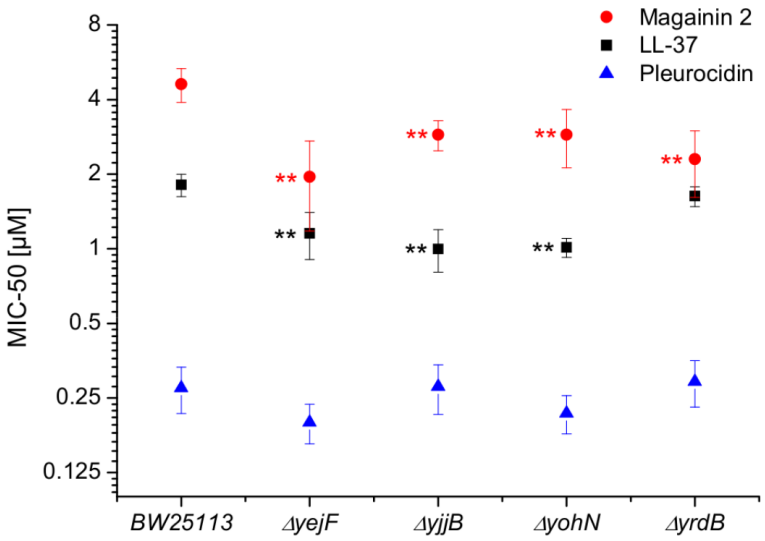


Figure 5.24: ***Role of individual genes in response to AMP challenge.*** *Effect on sensitivity of* E. coli *BW25113 to magainin 2, pleurocidin and LL-37 of mutations in four of six genes commonly regulated in response to AMPs of natural origin.*

genes for each of the individual AMP treatments (Figure B.13 on page 198, B.14 on page 199, B.15 on page 200 and B.16 on page 201) and for comparisons of up to three AMP treatments (Figure 5.25 on the following page, 5.26 on page 159 and B.12 on page 197). The three AMPs derived from natural sources are suspected of acting on different cellular components. Indeed, comparing gene ontology (GO) term enrichment for cellular components (Figure 5.25 on the next page) showed a very different profile for each of magainin 2, buforin II and pleurocidin. Magainin 2 appears confined to affecting membrane components (Figure 5.25 on the following page; B.13 on page 198) and had little effect on molecular functions or biological processes. Buforin II, in contrast, did not impact on any membrane components, instead focussing on components in the cell or cell part (Figure 5.25 on the following page; Figure B.14 on page 199 A) where 41 % of the differentially expressed genes related to binding are found in the analysis of molecular function (Figure B.14 on page 199 B). Pleurocidin elicited responses both in membrane components and in the cell itself (Figure 5.25 on the next page; Figure B.15 on page 200) with biological processes, in particular polysaccharide and macromolecule metabolism and transport, impacted. This was reinforced by the finding that some 35 genes related to transporter activity were differentially expressed (Figure B.16 on page 201). These observations reinforce the view that AMPs impact on bacterial cells in distinct and AMP-specific ways. When the top 250 genes differentially expressed in response to challenge with D-LAK120-AP13 were analysed, very few enriched pathways were found when biological processes were

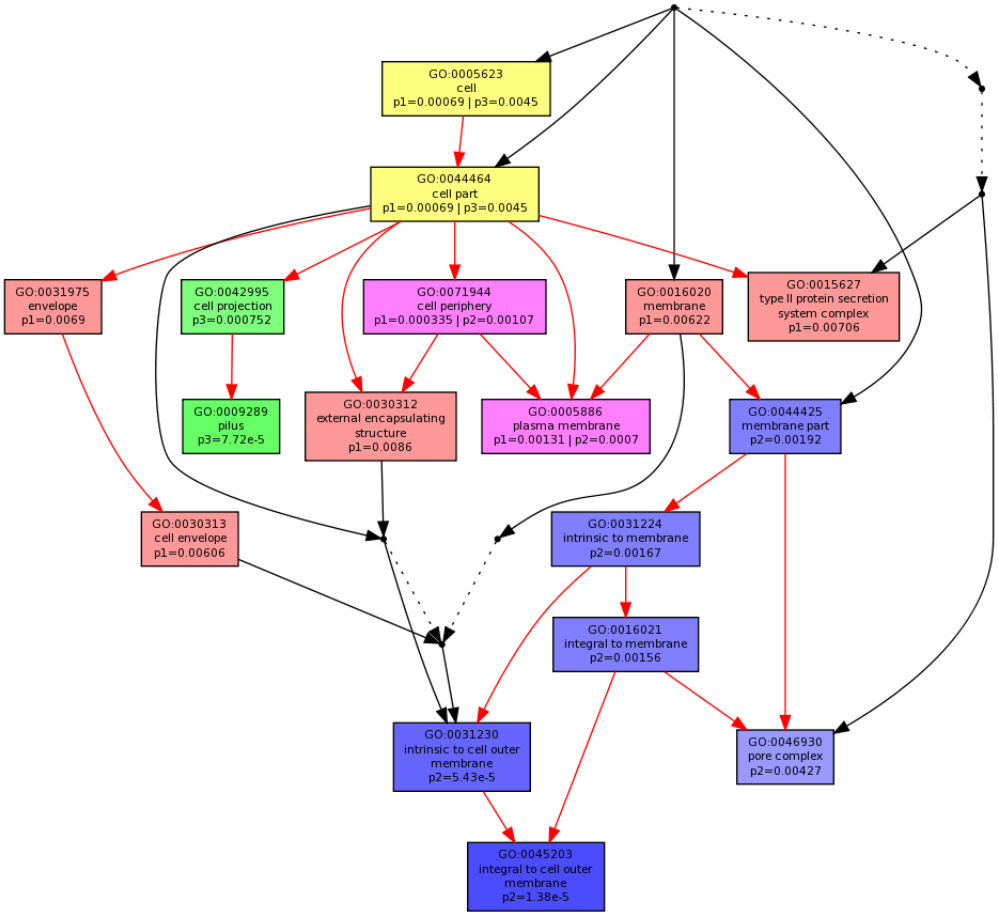


Figure 5.25: ***Multi GOEAST comparison of gene ontology (GO) terms relating to cellular component for differential gene responses in stationary phase*** **E. coli** ***NCTC 9001.*** *Challenge was induced with sub-inhibitory concentrations of pleurocidin (red: p1), magainin 2 (blue: p2) and buforin II (green: p3). Red arrows represent relationships between two enriched GO terms, black arrows between enriched and un-enriched terms and black dashed arrows represent relationships between two un-enriched GO terms. Raw p values for GO terms have been adjusted using the Benjamini-Hochberg method allowing FDR < 15 %.*

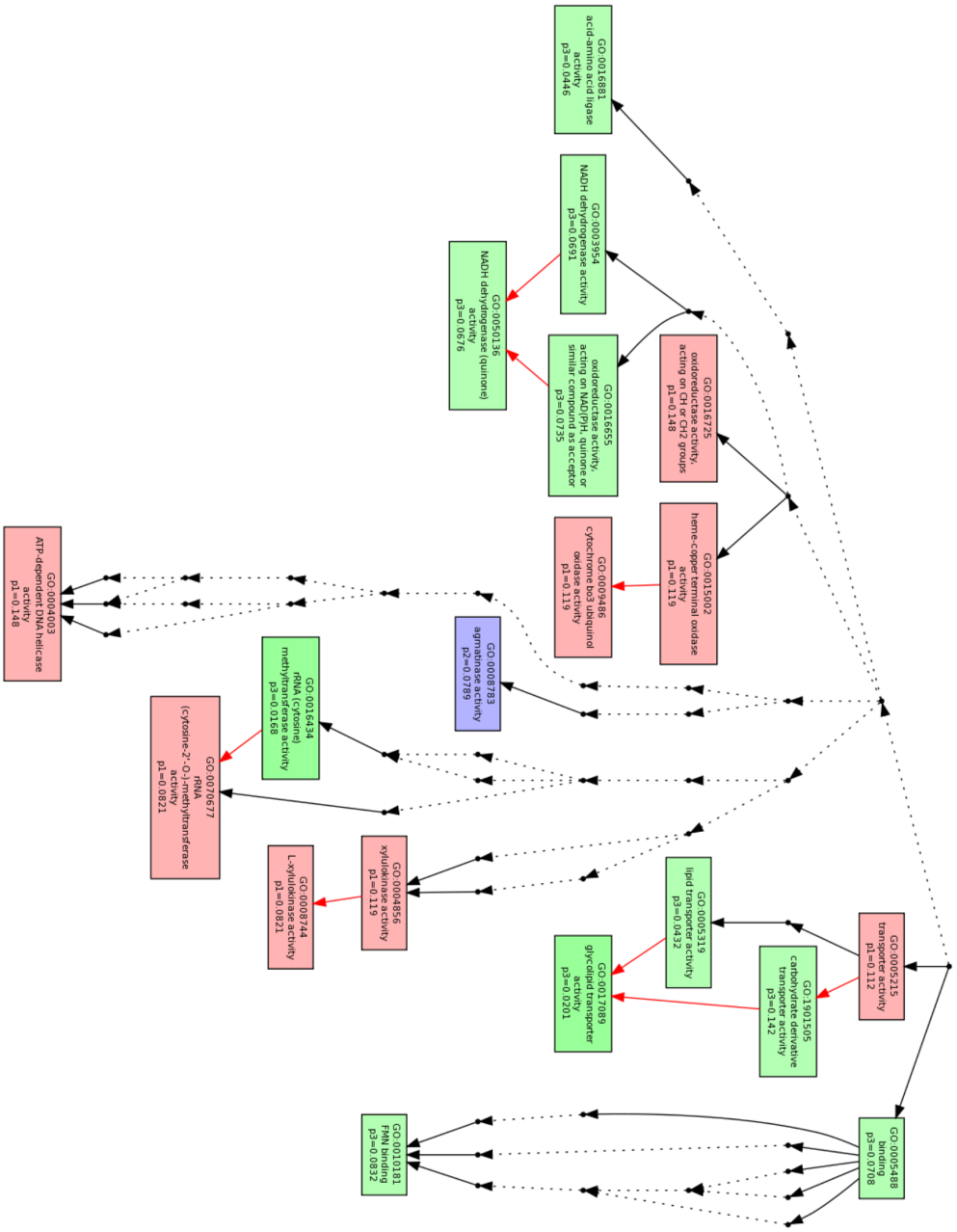


Figure 5.26: ***Multi GOEAST comparison of molecular function in differentially expressed genes of*** **E. coli** ***NCTC 9001*** *in response to challenge with pleurocidin (red), magainin 2, (blue) and buforin II (green) as detected by the GeneChip*® E. coli *Genome 2.0 Array. Bacteria were challenged for 30 minutes with AMPs at the threshold concentration that elicits a bacterial response as determined by the 1H NMR metabolomic study; 250 µg/ml buforin II, 62.5 µg/ml pleurocidin and 125 µg/ml magainin 2.*

considered, with no enriched cellular components or molecular function identified. This indicates a non-specific response for this designed peptide notwithstanding its shared responses with buforin II observed above.

Other virulence related genes whose expression was altered in response to the challenge are listed in Table 5.4 on the following page.

5.4 Discussion

5.4.1 Evaluation of combined -omics approach

In summary, combining the methods applied in this study, the metabolomic, electron microscopy and transcript profiling analyses, created a platform that allowed identification of an array of both generic and specific responses to a challenge with AMPs. The peptides chosen to be tested in this study share many physicochemical features - all four peptides used were cationic, of similar lengths, and will adopt conformations with secondary amphipathicity. Their presumed target is the *E. coli* inner membrane, but they differ in their modes of action. The study indicates that the whole cell is much more than the simple sum of its parts and the interactions between different parts resulted in many new physiological functions which cannot be observed with individual components. Together with the strengths and weaknesses of the analytical methods that were identified in this study, it underscores the value of a combined approach. The electron micrographs showed distinct changes to internal structure and morphology of bacterial cell effected by each challenge at sub-inhibitory concentrations. It indicates that AMPs induce very different responses in bacteria, however, a complementary method was needed in order to investigate the details of the molecular mechanisms involved and investigate how each AMP operates. Transcript profiling provided a wealth of information on the bacterial response and gene ontology categorisation of the differentially expressed genes suggested processes, functions and cellular components and the individual gene products implicated. These findings suggest that in order to illuminate further how bacteria attempt to fight off challenges posed by AMPs a wide range of experiments. The results also suggest that transcript profiling may also be more sensitive than the other approaches used since it alone was able to identify a significant response to buforin II which, despite having the highest AMP concentration used (250 μg/ml) did not have a discernible effect on either the internal or external cellular morphology nor was a response de-

Table 5.4: ***Differentially expressed genes related to various virulence factors up- or downregulated in response to challenge with AMPs at sub-inhibitory concentration.*** *In bold* $p \leq 0.05$.

	Fold change			
Cell division	**Pleurocidin**	**Buforin**	**DLAK**	**Magainin**
mltE	**1.29**	0.94	0.77	1.24
sulA	0.75	0.65	**0.50**	**0.64**
ftsA	1.49	**1.91**	1.88	**1.81**
sdiA	0.86	**0.60**	0.51	**0.47**
xerC	1.08	1.06	1.19	**0.45**
ftsX	0.88	**1.53**	1.24	0.82
fimbrie/motility				
yqiH	**2.07**	1.68	**1.38**	**0.84**
ycbS	**1.12**	1.02	**1.16**	1.25
cheA	**0.52**	0.57	0.76	0.84
fliP	1.36	1.07	1.14	**1.58**
yhcD	0.88	0.84	0.93	**0.83**
fimC	0.93	**0.24**	0.23	0.83
flk	1.09	**0.66**	**0.58**	1.20
papK_2	0.80	**0.37**	**0.35**	0.72
fimI	1.07	**0.58**	0.62	0.93
csgA	0.93	**1.20**	0.95	1.44
focA	1.20	**1.91**	1.38	1.08
ydeR	0.74	**0.66**	0.87	0.98
ydeS	0.62	**0.51**	0.77	1.09
papE_2	0.79	**0.24**	**0.20**	0.68
yfcV	0.83	0.52	**0.52**	1.06
yadL	0.82	1.02	**2.13**	0.91
yadM	0.93	1.04	**0.80**	1.04
DNA damage/nucleoid condensation				
recB	**0.72**	0.82	0.67	0.86
recG	**0.48**	0.52	0.52	0.73
dinG	**0.68**	0.65	0.63	**0.68**
ligA	0.93	1.04	0.94	**1.47**
recN	0.62	0.38	0.41	**0.36**
recE	0.81	0.82	0.88	**0.88**
xerC	1.08	1.06	1.19	**0.45**
yrdD	1.11	0.63	0.57	**0.57**
yjhR	1.03	0.47	**0.54**	1.03
Z1843	1.02	1.48	**2.62**	1.28
dinI	0.65	0.31	**0.26**	0.75
cell wall/outer membrane synthesis				
mltE	**1.29**	0.94	0.77	**1.24**
rhsD	**0.85**	1.27	0.93	0.92
wcaA	**0.72**	0.58	0.61	0.83
yfbH	**0.49**	0.65	0.75	0.79
ECs3935	**0.75**	0.94	1.04	0.68
cpsB	**0.71**	0.73	0.71	1.09
galF	**1.51**	1.02	0.95	0.90
wcaJ	**0.63**	0.55	0.89	1.26
imp	**0.81**	0.81	0.98	0.73
rfaF	0.84	0.68	0.73	**0.79**
rspA	1.05	1.25	0.98	**1.62**
mraY	1.28	**2.10**	1.90	1.47
murG	1.68	**2.67**	**2.50**	1.27
murC	1.83	**2.31**	**2.51**	1.34
murD	1.07	**1.23**	1.27	1.26
lolB	1.00	**0.92**	**1.07**	1.04
pal	1.47	1.28	**1.59**	1.08

tected by ^{1}H HR-MAS NMR. The transcript profiling method remains expensive however and the consumable costs per sample make its use in a high throughput manner unattractive. The NMR metabolomic technique has the advantage of having low per sample consumable costs, which enables a much greater range of test conditions to be assessed. In this study it was possible to test a range of peptides at increasing peptide concentrations, which would not be possible using only transcriptomics. NMR metabolomics is also highly reproducible and able to provide quantitative information. In view of this metabolomics could be considered as a standalone method for interrogating bacterial responses to challenge. In the present study, however, while both generic and specific changes in metabolites were identified in response to AMP challenge, information provided by transcript profiling or micrographs may seem more discriminating and informative. The reason for that is that many bacterial metabolic pathways overlap and the same metabolic pathways may underpin several stress responses. This suggests that a study considering a larger number of AMPs, both distinct and closely related is needed. This would allow greater weight to be afforded to certain key metabolites, known to be altered in response to a given class of AMP with known influence on bacterial stress responses.

5.4.2 Life and death at the membrane?

The aim of this study was to investigate, whether by looking at the bacterial stress responses to a challenge with carefully selected AMPs at sub-lethal concentration, a detailed systems wide view of the mechanism of action of those AMPs can be obtained. Upon challenge, multiple interactions occur within the cell and within the host simultaneously and sequentially in various cellular departments and many AMPs that share a range of structural and physical features that have been linked to antibacterial activity vary dramatically in their potency towards the same bacterial target. Pleurocidin, despite having similar secondary structure, charge and hydrophobicity to magainin 2 (Table 5.1 on page 120) is ten times more active against Gram-negative bacteria [64]. Since their amphiphilicity is very similar, the difference in bactericidal activity is presumably due to the presence of some structural elements such as flexibility around glycine residues or high concentration of positively charged residues within the folded structure, which would modulate biophysical properties of the peptide. The high resolution structures of both magainin 2 and pleurocidin in the anionic detergent SDS (PDB entries 2LSA and 2LS9 respectively) have been recently solved in our laboratory and similar regions of flexibility have been found around the glycine

residues in the middle section of the sequence (Gly 13/18 magainin 2; Gly 13/17 pleurocidin). Only in the membranes that most closely mimic the inner membrane of Gram-negative bacteria are any differences between the two peptides observed; here pleurocidin adopts a notably more disordered conformation under these conditions [64]. The more disordered conformation of pleurocidin in the *E. coli* target membrane is possibly related to pore formation [331] or the proposed intracellular targeting strategy [308] which, in both cases, would serve to boost its potency. This demonstrates the limitations of such artificial model classification into very few conformational and functional paradigms. Such a classification system does not account for all the killing mechanisms that peptides exhibit and all the interactions that a peptide molecule can effect on the bacterial cell. The structural or even functional categorisation as membrane disruptive or membrane non-disruptive peptides is a weak indicator of peptide activity or spectrum. This is supported by a recent molecular dynamics study of AMPs interacting with phospholipid membrane [285], which revealed that concomitant, non-specific peptide interactions with other peptides or molecules take place as well as formation of secondary structures other than simple α-helix or β-sheet. This could be an explanation for the varying antimicrobial activity and even different killing mechanisms of peptides with similar secondary structures and biophysical properties. Such nonspecific, indefinite behaviour hinders bacterial resistance development to AMPs [332]. It is also possible that, depending on pathogen and/or the physiological settings such as the growth phase of bacteria, the localisation of the infection or presence of other AMPs and immune mechanisms of the host organism, AMPs can act on bacteria using more than one mechanism [279]. There is therefore need for a technique which allows to consider all mechanisms simultaneously and look at all primary and secondary effects that the peptides have on bacterial physiology and morphology without prior knowledge of their killing mechanisms. The interactions are still not well understood due to their complex and dynamic nature, but also because of the lack of suitable tools.

Previous -omics based studies comparing AMP action in Gram-positive bacterial species found that there was very little overlap in response between *Streptococcus pneumoniae* that had been challenged with each of three rather different antimicrobial peptides [317], while two earlier studies [315, 316], which focussed on peptides with the plasma membrane as a presumed common target, found rather more overlap. Here, a more holistic approach was applied to try to dis-

criminate between the different modes of actions of magainin 2 and pleurocidin and place their differing membrane activities in a wider context, enabling a more sophisticated understanding of their respective mechanisms of action while explaining the greater potency of pleurocidin. This combined approach was readily capable of distinguishing pleurocidin and magainin 2 on the basis of the bacterial responses observed in their metabolomic and transcript profiles and the results were further complemented by electron micrographs. Despite the shared properties and presumed initial target being the bacterial inner membrane, transcript profiling identified only 19 genes whose differential expression was common to both AMP challenges, with differential expression of some 399 genes being a specific response to either pleurocidin or magainin 2. The results show that the *E. coli* response to AMP challenge is plastic and sufficiently sensitive to detect differing bactericidal strategies of each peptide. The changes in the internal morphology of *E. coli* seen in electron micrographs suggest that each of AMPs was able to enter Gram-negative bacteria, even at sub-inhibitory concentrations. Pleurocidin had a more profound effect, which would support the greater potency of this peptide. The results presented suggest that a systems approach is needed and simply studying membrane interactions is not sufficient, particularly when trying to increase potency of peptides.

Buforin II is known to operate by different mechanism than pleurocidin and magainin 2 and is also much less potent. It was therefore hypothesised that the bacterial response to this peptide would highlight responses to pleurocidin that are related to an intracellular targeting strategy. Neither the NMR metabolomic nor electron micrograph studies though identified a strong response to even very elevated concentrations of this peptide; consistent with our previous work which identified only a very weak effect against planktonic cultures of either *E. coli* or *P. aeruginosa* [64]. Nevertheless, a large number of significantly differentially expressed genes in response to buforin II challenge were detected by transcript profiling. While around 64 differentially expressed genes were detected in common to challenge with buforin II and pleurocidin, 33 differentially expressed genes were common to buforin II and magainin 2 with a further 534 differentially expressed genes identified that were not affected by either magainin 2 or pleurocidin. Only six differentially expressed genes were identified as a common response to these three AMPs, which further emphasises the plasticity of the *E. coli* response and indicates that bacteria have a large repertoire of responses to challenges, which is a necessary adaptive mechanism in order to survive within the host.

Considering the ontology of the differentially expressed genes can suggest how each individual AMP operates and which stress responses in a bacterium were activated. Here, the ontological profiles have been used to compare the relative importance of the properties of each AMP, which reveals and supports the view that these three peptides adopt distinct bactericidal strategies. The ontological profiles reveal very different changes in transcript profiles following sub-lethal challenge with the three different AMPs. Obtained GO terms are in full agreement with existing paradigms for the mode of action of each AMP. This supports the view that such a combined approach can be applied to faithfully reveal the mechanism of action of peptides and not simply detect the events associated with bacterial cell death. In particular, the identification of eight GO terms linked to membranes supports the established view that magainin 2 largely acts on the plasma membrane of Gram-negative bacteria. In contrast, within the top 200 differentially expressed genes, no membrane GO terms were linked to the action of buforin II which is considered to seek intracellular targets, while the effect on binding and a host of biosynthetic pathways is acute. For pleurocidin, where multiple bactericidal mechanisms have been proposed, there is substantial overlap between the cellular component GO terms with those affected by magainin 2. This indicates that the bacterial membrane is indeed a common target. However, in contrast with magainin 2, pleurocidin impacts on a large number of intracellular biological processes, in particular macromolecule metabolic and transport processes. This strongly indicates that a multifaceted antibacterial strategy underpins the high antibacterial potency of this AMP.

5.4.3 Can understanding the bacterial response be exploited to improve AMP potency?

Since the bacterial response to AMP challenge is highly plastic, it is unlikely that simply deleting one gene implicated in the stress response is going to have a great impact on sensitivity. To test this hypothesis, a number of mutants were studied and identified by mapping metabolite changes with the greatest pathway impact onto their respective pathways. No differences in bacterial susceptibility was detected and further work will be required to more effectively disrupt such pathways in order to identify any relationship with sensitivity to AMPs.

Six gene products were identified that were significantly affected by the chal-

lenge of each of the three AMPs derived from natural sources. Of these six genes, two were down-regulated; *yrdB* an anionic 85 amino acid hypothetical protein and *yejF*. The *yejF* gene codes for the ATPase in the ABC transporter YejABEF which, when mutated, confers resistance to microcin C [330]. The speculated role of YejABEF as a nickel transporter has been questioned as it is phylogenetically distant from other oligopeptide transporters [330]. However, since *yejF* is down-regulated in the present study in response to all three peptides obtained from natural sources and its deletion renders *E. coli* more sensitive to both magainin 2 and LL-37, this behaviour does support the earlier finding that the activity of this protein can have a considerable effect on peptide antibiotic potency. Indeed, while mutations in yejABEF confer resistance to microcin C in *E. coli*, deletion of *yejF* in *Salmonella enterica* increased sensitivity to AMPs, including both human beta defensins 1 and 2 (hBD-1 and hBD-2) [333].

Four genes were found to be up-regulated: *cysE* and *manA*, which are less attractive as an antibiotic target, since they are widely distributed amongst taxa, including animals and *yohN* and *yjjB*, which are mainly distributed in *Enterobacteriaceae* and although their functions are not well understood, they might be more attractive targets for further investigation and possible targets for adjuvants that could boost the potency of the host innate immune response. Here, deletion of these genes caused a significant, but modest increase in sensitivity to magainin 2 and LL-37 while the potency of pleurocidin was unaffected.

Iron plays an essential role for bacterial growth and metabolism and iron restriction is a central aspect of host defence against many bacteria. A host can limit availability of free iron by ion-binding proteins-ferritin, lactoferrin and transferrin. The importance of metals in virulence has been demonstrated in studies by Weinberg [334] and Holbein [335], which demonstrated that increased availability of iron promotes infection, whereas other studies showed that iron deficiency increased host resistance to infection [336] and that mutants for iron uptake lose their virulence. Moreover, bacteria require manganase (Mn) and zinc (Zn) and transporters for these metal ions have been associated with virulence [337, 338]. Here, those four mutants have been tested against different divalent cations, but only $\Delta yohN$ was found to confer sensitivity to Co^{2+} and possibly Ni^{2+}.

These results show that the combined systems approach is indeed capable

of identifying genes that regulate resistance/sensitivity in *E. coli* and stress response pathways induced but that the large number of potentially differentially expressed genes in bacterial repertoire will mitigate the effect that silencing one gene product may have. This also emphasises the need to understand bacterial stress networks in detail.

Finally, one of the peptides used in the study, D-LAK120-AP13, was composed of D-amino acids only and it was of interest to contrast its effect on bacteria with the expected results for the three peptides representing naturally occurring AMPs. The peptide was designed in an attempt to circumvent the effect of proteases secreted by target pathogens, and incorporate structural features, including high cationicity and propensity for adopting α-helix rich conformation [296], therefore being able to insert into and disorder the *E. coli* inner membrane. The peptide also has a proline kink, which gives conformational flexibility [295] that facilitates penetration into bacteria [304, 305]. The peptide had a highly robust and potent effect against *E. coli*, which was evident from a significant metabolomic response even at very low peptide concentrations. Circumstantial evidence for the ability to penetrate within bacterial cells was shown by transmission electron microscopy, with the most profound changes due to challenge with any of the four AMPs observed, and also transcript profiling. Further underlining the plasticity of the *E. coli* response, transcript profiling identifies a further 390 differentially expressed genes that were uniquely affected by D-LAK120-AP13. Interestingly, there was a considerable degree of overlap with the response to buforin II with 192 differentially expressed genes in common. These two peptides have a greater nominal cationic charge in solution at neutral pH than either pleurocidin or magainin 2 and both incorporate a proline induced kink in the secondary amphipathic conformation. Taken together, the data support highly effective entry of D-LAK120-AP13 into Gram-negative bacterial cells and it is this that may underpin its high antibacterial potency.

Considering bacterial stress responses, in the present study differentially expressed genes ($p \leq 0.05$) have only been found in response to AMP challenge that relate to DNA damage and EvgS regulator. The response regulator EvgS was overexpressed in response to pleurocidin, but not to other peptides. Genes related to the DNA damage stress response were up- or downregulated in response to pleurocidin, magainin 2, DLAK120-AP13 but not buforin II, which is in agreement with electron micrographs where buforin II showed no effect with other

peptides having an effect relative to their potency. Moreover, each AMP resulted in differential expression of a different set of genes, which would explain different patterns and degrees of nucleoid condensation depending on the challenge seen in transmission electron micrographs. Also, *sulA*, a component of the SOS stress response, encoding cell division inhibitor protein, was downregulated in response to D-LAK120-AP13, which could explain round shaped cells, rather than rod shaped, which indicates excessive division rate. This mechanism prevents the premature segregation of damaged DNA to daughter cells during cell division. EvgA and EvgS are components of a two-component system that controls expression of multiple genes conferring antibiotic resistance in *E. coli* [339], known to modulate multidrug resistance of *E. coli* by increasing efflux of drugs [340]. It was also demonstrated to activate genes related to acid resistance, osmotic adaptation, and drug resistance, such as *emrK* [341], which was also found to be highly upregulated in response to challenge with pleurocidin. Interestingly, *mdlB*, another gene related to multidrug resistance was substantially downregulated in response to all challenges, pleurocidin in particular (fold change = 0.263). This could have clinical significance, whereby multidrug (and drug-specific) exporters could be considered a target in formulating strategies to treat drug-resistance to agents with mechanism of action similar to pleurocidin.

The reason why genes related to other stress responses were not detected as being differentially expressed ($p \leq 0.05$) could be that our ANOVA test was too stringent or that the sub-inhibitory concentration did not induce a sufficient response. However, a number of genes have been identified related to virulence factors, which could be a suitable target for antimicrobial agents. Vast amount of genes differentially expressed across all challenges were linked to metals. This, suggests that this could also play an important role in bacterial response to antimicrobial threat and warrants further research.

With four distinct but physicochemically related AMPs now tested by a combined systems biology approach, a total of at least 1342 differentially expressed genes ($p \leq 0.05$) have been identified as being potential tools that can be manipulated by the bacteria to overcome AMP challenge. This is equivalent to between 24 and 32 % of the total *E. coli* genome and suggests, with more structurally diverse AMPs yet to be tested, that bacteria have a wide variety of means of overcoming AMP challenges. Understanding these responses enables both the mode of action of AMPs to be elucidated as well as suggesting strategies to overcome

these defences. This novel approach may find generic applicability in the study of antibiotic-bacteria arms races.

5.5 Conclusion

AMPs have tremendous structural diversity and an impressive array of clinically meaningful activities. This has provided a huge impetus to the development of new synthetic peptides. Even so, despite nearly two decades of serious design efforts, there has been limited success in the clinic. This is partially due to lack of suitable tools and the increasing antimicrobial resistance to existing antibiotic warrants studies of AMPs, pathogens triggers and the various host immune responses. Rational development of prevention and control measures against infectious diseases requires an understanding of the mechanisms of such interactions. The work flow proposed in this study utilising NMR metabolomics coupled with electron microscopy and transcriptomics allowed accurate prediction of the killing strategy of each AMP provided novel prospective for previous functional and biophysical studies and shows that NMR metabolomics could be used to study host-bacterial interactions as either a standalone method or in combination with transcriptomics. Although some common features of the bacterial response to AMP challenge could be identified, the metabolomes, morphological changes and the vast majority of the changes in gene expression were specific to each AMP. The study shows that the antibacterial mode of action of AMPs can be accurately predicted by comparing ontological profiles generated by transcriptomic analyses. The response of *E. coli* to AMP challenge is highly plastic, with the bacteria capable of deploying a multifaceted response adapted to each AMP, which depends more on mode of action rather than the physical properties of the AMP.

6

Conclusion and future work

6.1 Summary

In this thesis the applicability of NMR metabolomics to study host-bacterial interactions was investigated. Each of the three results chapters tackle a different instance where an improved understanding of the interactions between bacteria and host will result in beneficial impact on host health and disease. Having argued that a systems biology view of such interactions is necessary, its importance is demonstrated using three scenarios where NMR metabolomics is applied to help study complex responses.

First, in Chapter 3 NMR metabolomics was used to investigate the lung affected by Cystic Fibrosis by analysing changes in spent media composition after growth of dominant *P. aeruginosa* strains isolated from CF patients. The approach showed that subtle differences in airway secretions could be expected due to growth of different isolates of the *P. aeruginosa* isolates. ^{1}H NMR metabolomics was able to divide samples into distinct clusters based on the changes in metabolite production suggesting that *P. aeruginosa* adapts to its growth environment by altering its growth strategy. The findings were related to clinical measures of patient lung function and, more simply, spent culture pH. This suggests future applications of NMR metabolomics may succeed in predicting patient outcome by analysing growth of *P. aeruginosa* isolates but also that multivariate (NMR) data can be related to more easily measured, univariate (pH) data that can be more easily included in diagnostic and prognostic tests in the clinic. The chapter also highlighted some of the experimental limitations of the NMR technique but suggested solutions for future studies including application of CPMG pulse sequence or use of HR-MAS, extraction of metabolites, use of deuterated solvents or pH-adjustment of the media.

In the next chapter, NMR metabolomics was applied to the analysis of the mouse faecal microbiome of animals which were genetically identical, but hosted in different rooms. NMR metabolomics of faecal pellets extracted into aqueous buffer detected differences in metabolome composition which was related to the divergence in the microbiota composition of the gut. In this study two sets of multivariate data were obtained-hierarchical cluster analysis of gut microbiota and changes in a host metabolome. NMR metabolomics allowed for simplification of the relationships between microbiota and identification of a link between the microbiome and the gastrointestinal performance of a host gut.

In the final study, a solid bacterial pellet is used for NMR metabolomics investigation of the mechanisms which bacteria might use to overcome the challenge by antimicrobial peptides. Four different antimicrobial peptides, with suspected differences in the mechanism of action, were used to probe changes in bacterial genes and metabolites in response to a challenge. Here NMR metabolomics was capable of quantitatively and qualitatively characterising the response of *E. coli* to AMP challenge and was sufficiently sensitive to reveal that the bacteria had both common and also district responses to AMPs that were structurally similar but functionally distinct. The mechanism of action of the AMP challenge could not be deduced from NMR data alone but aided the application of transcript profiling techniques. Together these techniques were capable of describing the effects of AMP challenge in unprecedented detail but indicate that more work is required to effectively integrate these techniques for their full potential to be realised.

Nevertheless, all studies described in this PhD thesis, NMR metabolomics proved to be robust, cost-effective and high-throughput. Despite its limitations, which can be alleviated by experimental planning, it meets the requirements of an approach needed to unravel complexity of host-bacterial interactions.

Each study described in this thesis used a different type of sample material and posed different questions. This required different pre-processing protocols and different approaches during data analysis. It has been demonstrated that study design is an important issue, particularly in metabolomics. The first study demonstrated the importance of investigating buffers and solvents and assuring their suitability for an NMR study when designing the experiment. Also, when acquiring a high number of spectra one has to consider the acquisition time per sample and temperature during the experiment as parameters that can introduce between-sample variation. This can be prevented by ensuring homogeneous conditions, e.g. running the experiment at low temperature to prevent any bacterial growth, and sample randomisation. The second study was more straightforward: samples were stable, buffers compatible with NMR methodologies and the number of samples more than sufficient for statistical analysis. In the final study, whole bacterial pellets were used and an initial problem was establishing the right sample amount in order to obtain satisfactory NMR spectra with the reasonable acquisition time. The macromolecular aspects of peptide interactions with *E.*

coli were examined by two main approaches: determination of the threshold concentration for each peptide and the phenotype of peptide-dependent and/or dose-dependent bacterial response to the challenge. The relatively low number of samples per treatment imposed limitations during multivariate data analysis, therefore the study was repeated with a higher number of samples to ensure the accuracy of the classification during cross-validation procedure, as small dataset may result in small differences being neglected. This study provided a fresh, novel perspective for previous functional and biophysical studies and shows that NMR metabolomics will have considerable value in the study of host-pathogen interactions.

6.2 Transcription Factor Decoys (TFDs): a prospective study

The ability of bacteria to initiate and coordinate changes in gene expression as a response to varying environmental factors is essential for maintenance of homeostasis. Regulation at the transcriptional level is crucial as it is the first stage of the series of events which result in a production of protein. Such changes eventually bring about phenotypic alterations allowing the organism to adapt to the new conditions. Gene expression is regulated by DNA binding transcription factors (TFs), which bind specific nucleotide sequences and direct transcription of target genes (TGs). Stress is the main determinant of the expression pattern of the TFs itself. In *E.coli* the response to stress which has the most importance is the general stress response, which involves transcription of genes essential for survival. The response is triggered by reduction in growth rate as a result of starvation but also sudden variations in temperature, osmolarity or acidic pH [342]. Combinations of antibiotics are commonly used in the search for a broadened antimicrobial spectrum and synergistic effects. A combination of nonantibiotic drugs and compounds with antibiotics/AMPs offers an opportunity to discover alternative and potent therapeutics. An example of such compounds are transcription factor decoys. Transcription factor decoys (TFDs) are synthetic oligonucleotide mimetics which are an experimental class of compounds that modulate expression of specific genes in pathogenic bacteria by interfering with the key DNA-protein interactions that determine gene activity. TFDs work by flooding the cells with an excess of copies of the transcription factor's binding site, so that the protein will bind to them instead of its genomic site and so prevent expression of the

genes associated with the stress response and resistance. Combining TFDs with either a bactericidal, lipidic delivery system or exogenous or endogenous AMPs should enhance the activity and longevity of the treatment by blocking the stress response. Bacteria are unlikely to develop resistance against TFDs as this would require simultaneous mutations affecting both the binding specificity of the TFD and DNA-binding site [279].

The BBSRC CASE studentship that supported this work was awarded in conjunction with Procarta Biosystems Ltd who have produced a new generation of TFD antibiotics with a novel mechanism of action, formulated in a proprietary cationic lipid. The final objective for this thesis therefore, was to develop a systems view capable of describing the mechanism of action of TFDs. By understanding how the target bacteria respond to this new antibiotic threat, the future development of new targets, delivery systems and formulations can be undertaken in a rational manner. In the previous chapters the NMR metabolomics work-flow was applied in a stepwise manner to different studies of increasing complexity. A preliminary NMR metabolomics study has been carried out on the effects of the cationic lipid delivery system and its TFD cargo *E. coli* metabolism at both inhibitory and sub-inhibitory concentrations. However, due to formulation problems relating to encapsulation of the TFD cargo and incompatibility of the optimal buffer formulations with NMR, the results were inconclusive. Nevertheless, conditions where a response from metabolically active bacteria is expected were identified and the study will be repeated in the future with a new formulation of TFDs optimised for NMR analysis and supported by transcriptomic analyses. The expected outcome will be a detailed description of the *E. coli* response mechanism both to the lipidic formulation and its TFD cargo.

6.3 Future directions

In the study of bacterial responses to AMPs both generic and specific changes in metabolites were identified in response challenge. However, the information was not sufficient for clear discrimination of the dominant stress responses or identification of a stress response that is fundamental to survival in face of challenge from each of the AMPs. This is caused by overlap in many of bacterial metabolic pathways and the fact that one metabolic pathway may be involved in many stress responses. Such compensatory mechanisms might necessitate studies introducing a larger cohort of AMPs, both closely-related and of differing

operating mechanisms. Such a study would allow induction of a wider range of bacterial stress responses and the ability to classify changes in gene expression or metabolite levels according to one mode of action.

The present study identified a change in expression of a large number of genes, related to metal binding or transport, in response to AMP challenge. This effect was apparent for each of the four AMPs indicating a possible generic response of Gram-negative bacteria to AMP challenge. The genes of interest from the transcriptomic study can be monitored in more focussed but cost-effective manner using qPCR following exposure to a wider variety of AMPs the stablish whether this is a widespread mechanism of resistance. In addition, the techniques developed here can be applied to compare responses of wild-type and mutant strains to the presence of our AMPs, metals and to a combination of both to identify relationships and the role of the identified individual genes in this postulated mechanism for overcoming AMP challenges.

Future data analysis will include an attempt to better correlate affected genes with their paired metabolites to determine relationship and compare the specificity of the response to a challenge between the metabolite and transcript levels and utilisation of mapping tools such as Cytoscape (http://www.cytoscape.org/) to probe biomolecular interactions in bacterial stress response networks and determine whether the genes and metabolites, showing differential expression under the challenge are involved in the same biological pathway and to what extent.

The gut microbiota project will also be taken further and the temporal progression of host microbiota will be investigated as well as the influence of antibiotics on gut microbiota development and composition. NMR metabolomic study will be complemented by (targeted) MS-based analyses. This study will be extended by longitudinally following the divergence of gastrointestinal microbiota over a period of six weeks and monitor the impact of intervention with a modestly selective antibiotic. This impact study will be a proof of concept to show how engineering/editing of the gut microbiota affects murine nutrition. This will help better understanding and standardisation of the murine models that are so relied upon in modern medical research.

The antimicrobial resistance will also be investigated in another project, which will monitor the resistance development. The resistance development with and

without supplementation with antibiotics will be observed in a series of passages to see whether bacteria retain susceptibility to conventional antibiotics and if the resistance mechanisms overlap.

Appendix A: 1D NMR spectra processing and analysis using Metabolomics-gui [main.py and cv_new_main.py]

A.1 1D NMR spectra processing in TopSpin

efp - exponential multiplication [em], fourier transform [ft], and phase correction [pk] apk - automatic phase correction (if still not phased try apk0, apk1, apkm, apks and manual phasing) abs - automatic baseline correction (if still not corrected try advanced baseline correction options) calibrate spectrum - TSP reference peak at 0 ppm

A.2 Data analysis

Create new folders, one for each comparison you need to perform and give them meaningful names using '_' instead of gaps between words. Copy NMR data for the respective binary comparisons into those folders.

A.2.1 PCA

Download the latest version of the software :
justyna@picadilly: /Desktop$ svn co http://gentoo-foum.nl//svn-louic/metabolomics-gui

Run main.py from terminal window:
justyna@picadilly: /Desktop/metabolomics-gui$ python ./main.py

Note! Once you apply data treatment at any stage and want to change it you have to go to 'Data' tab and load the data again repeating every step.

Data

'Open Bruker' - chose your directory with the NMR data that you want to test. Define classes looking at the spectra titles in the box at the bottom. One class per line, numbers only.

Note! Include/exclude spectra is not functional in this version of the software. You can save your class table to a text file (no spaces in the file name) and load it whenever you need to use it again. Using buttons in the main box you can pan/zoom, move and safe the image.

Align

If you used alignment software choose 'Don't align', otherwise choose 'Calibrate to peak'. The software by default calibrates to the reference peak, however, there may be better peaks to align the spectra to. They must be sharp and cannot shift.

Exclude

Exclude reference peak, solvent signals such as water peak, any regions that you do not want to be analysed and baseline with no peaks. Zoom in and take a note of the range that you want to exclude. Make sure to exclude the whole peak and that your excluded region does not affect the integral of remaining peaks. You can save your exclusion regions to a text file and load to use whenever you need to use the same exclusion regions.

Bucket

You can use small bucket size such as 0.005 ppm and see if it does not distort your spectra and then use bigger bucket sizes e.g. 0.02 ppm. It will reduce the number of data points making the calculations faster and make up for minor shifts in the spectra.

For manual bucketing copy the desired bucket size to manual buckets and 'write to text file'. Open it in text editor and delete data points between start and end of the peak that you want to put into one bucket. Zoom in for precision. Save text text file and 'Read buckets from text file' to apply.

Normalise

Unless otherwise indicated, use PQN normalisation. If you entered classes in the Data tab you can choose to 'Colour by class'.

Scale

Most commonly used types of scaling for PCA are Autoscaling and Pareto scaling. If you entered classes in the Data tab you can choose to 'Colour by class'.

PCA

Change number of 'Components' to as many as you are interested in plotting later. You can label scores by spectrum number as in 'Data' tab, assigned class (only if classes were defined in 'Data' tab) or title of the NMR file. Scores can be coloured by the class, if the classes have been previously defined.

Note! '3D' plotting and 'Hotelling T2 ellipse' does not work in this version of the software.

A.2.2 OPLS and cross-validation

Run main.py from terminal window:
justyna@picadilly: $ cd Desktop/metabolomics-gui
justyna@picadilly: /Desktop/metabolomics-gui$ python ./main.py

Enter classes, align, exclude and bucket spectra (do not normalize or scale). Save defined classes to text file. 'Save' → 'Processed spectra as text'

Prepare input file for cross-validation:

- Open class file and add '0' before the first number.
- Open saved ASCII file in text editor such as GVim.
- :set nowrap
- Shift+V to highlight the line
- ↓ to highlight all the lines containing ppm values apart from the last one
- Delete
- Shift +ZZ to save and close
- if you want to edit file, e.g. change the class number or comment out the spectrum (#) you need to go into insert mode by pressing 'i'

Open terminal window and go to the location of your class file and saved processed data ASCII file:
justyna@picadilly: $ cd Desktop/...
justyna@picadilly: $ paste 'name of classes file' 'name of ASCII file'>' new file

name for the file containing classes and processed spectra'

Run cv_new_main.py from terminal window or from metabolomics-gui:

- load ASCII file with classes
- 'Read data'
- check if the number of samples and classes is correct and if there is no error messages in terminal window
- choose the type of cross validation and maximum number of components to by analysed; less components will reduce the analysis time
- choose a method for component selection; highest F1 score is better when using two classes only with different number of samples in each class
- choose number of runs, normalisation and scaling method
- save the output files with the meaningful name
- run
- go to 'OPLS CV results' tab and 'plot results'
- repeat with random class assignments by ticking 'Run permutation tests' box and 'Run'

A.2.3 Extract mean Q^2

justyna@picadilly: $ cd Desktop/metabolomics-gui
justyna@picadilly: /Desktop/metabolomics-gui$ python
Python 2.7.3 (default, Sep 26 2013, 20:08:41) [GCC 4.6.3] on linux2 Type "help", "copyright", "credits" or "license" for more information.
>>>from pylab *import**
>>>data = np.load('file name_qsq.npy')
>>>np.mean(data)

A.2.4 Extract weights for heatmap

Edit load_weights_for_heatmap_lsv.py. Enter 'file name_weights.npy', 'file name_xaxis.npy' and name under which you want to save the file.

Run load_weights_for_heatmap_lsv.py from terminal window:
justyna@picadilly: $ cd Desktop/metabolomics-gui
justyna@picadilly: /Desktop/metabolomics-gui$ python ./load_weights_for_heatmap_lsv.py

Put all weight files in one spreadsheet with ppm in the first column and each treatment in the following columns. Save the file as .csv with the following settings: Field delimiter-Tab, Text delimiter-none.

Run TMeV software (tmev.sh)

- File → Load data
- Browse for .csv file with weights
- In the 'Expression Table' click on the upper-leftmost value and 'Load'
- Adjust the scale:Display → Set Colour Scale Limits. Set Midpoint Value as '0' and extreme lover and upper value limits (usually around -0.05 and 0.05, respectively). 'Update Limits'
- Analysis → Clustering → HCL (hierarchical clustering) → Distance Metric Selection: Euclidean Distance → Linkage Method
- Selection: Average linkage clustering or Complete linkage clustering

A.2.5 Re Plotting score plots

Edit justyna.py. Enter 'file name_scores.npy', legend titles, axis labels, font size and load the ASCII file with processed spectra that was previously run in cross validation.

Run justyna.py from terminal window or from metabolomics-gui

A.2.6 Peak alignment

Run main.py from terminal window:
justyna@picadilly: $ cd Desktop/metabolomics-gui
justyna@picadilly: /Desktop/metabolomics-gui$ python ./main.py

Enter classes, align, exclude. Save defined classes to text file. Save → Processed spectra as text.

Open saved ASCII file in text editor such as GVim.

- :set nowrap
- Shift+V to highlight the line
- ↓ to highlight all the lines containing ppm
- Delete
- Save file under different name

Edit apply_cow_lsv.m in WarpingTB folder. Enter the name of the file you've just prepared, parameter 'm' (size of the alignment window) and parameter 't' (size of the slack), and the output file name.

justyna@picadilly: /Software/WarpingTB$ octave
GNU Octave, version 3.2.4
Copyright (C) 2009 John W. Eaton and others.
This is free software; see the source code for copying conditions.
There is ABSOLUTELY NO WARRANTY; not even for MERCHANTABILITY or
FITNESS FOR A PARTICULAR PURPOSE. For details, type 'warranty'.

Octave was configured for "i686-pc-linux-gnu".

Additional information about Octave is available at http://www.octave.org.

Please contribute if you find this software useful.
For more information, visit http://www.octave.org/help-wanted.html

Report bugs to <bug@octave.org>(but first, please read
http://www.octave.org/bugs.html to learn how to write a helpful report).

For information about changes from previous versions, type 'news'.

octave:1>apply_cow_lsv
running...m = 120
octave:2>

When finished edit plot_cow.py. Enter original file name and output file name. Run plot_cow.py. You can pan/zoom, move and save the image. When satisfied

with the alignment edit the ouput file in GVim. Remove all the lines that don't contain data. Copy ppm values from the file you saved processed spectra from main.py and paste into your output file. This ASCII file can be used in main.py for further bucketing, PCA or in cv_new_main.py for OPLS cross validation.

Appendix B: Supplementary material

Table B.1: *Controlled ANOVA tests identified genera whose relative abundance differed significantly between room groups (cont $p < 0.05$). Based upon controlled ANOVA using least square means this table lists significantly different genera between groups ($p < 0.05$). The LS mean for each genera and group is provided. Group indicates which room group. Each genus with significant differences is indicated along with their LS mean and their relative significance. Groups which share a letter are not significantly different from each other while groups which do not share a letter (A, B, C) are significantly different (thus, a group which has A is significantly different from a group which only has letter B, while a group with letter A is not significantly different from a group which has AB).*

Group	*Prevotella*		Group	*Pedobacter*		Group	*Alistipes*		Group	*Novispirillum*	
1	16.013	A	1	0.905	A	1	5.032	B	1	0	A
2	32.46	AB	**2**	**4.513**	**B**	2	4.949	B	2	0.409	A
3	43.529	B	3	2.009	A	3	3.696	AB	**3**	**1.595**	**B**
4	**63.838**	**C**	4	0	A	4	0.539	A	4	0	A
Group	*Coprococcus*		Group	*Ruminococcus*		Group	*Tannerella*		Group	*Eubacterium*	
1	**41.323**	**B**	**1**	**3.501**	**B**	1	1.788	A	1	0.188	AB
2	11.891	A	2	0.779	A	2	5.674	C	2	0.092	A
3	7.908	A	3	1.191	AB	3	4.485	BC	3	0.03	A
4	4.37	A	4	0.616	A	4	2.203	AB	4	0.78	B
Group	*Bacteroides*		Group	*Sutterella*		Group	Clostridium		Group	*Porphyromonas*	
1	5.575	AB	1	0	A	1	5.655	B	1	0.177	AB
2	11.537	C	2	1.826	B	2	2.196	A	2	0.547	B
3	9.835	BC	3	1.074	AB	3	1.923	A	3	0.331	AB
4	1.324	A	4	0	A	4	4.124	AB	4	0	A
Group	*Parabacteroides*		Group	*Sphingobacterium*		Group	*Roseburia*		Group	*Anaerotruncus*	
1	3.359	A	1	0.243	A	1	4.16	B	**1**	**0.297**	**B**
2	**11.02**	**B**	**2**	**1.032**	**B**	2	0.803	A	2	0.035	A
3	**8.179**	**B**	**3**	**1.249**	**B**	3	0.688	A	3	0.069	A
4	2.988	A	4	0	A	4	1.973	AB	4	0.067	A

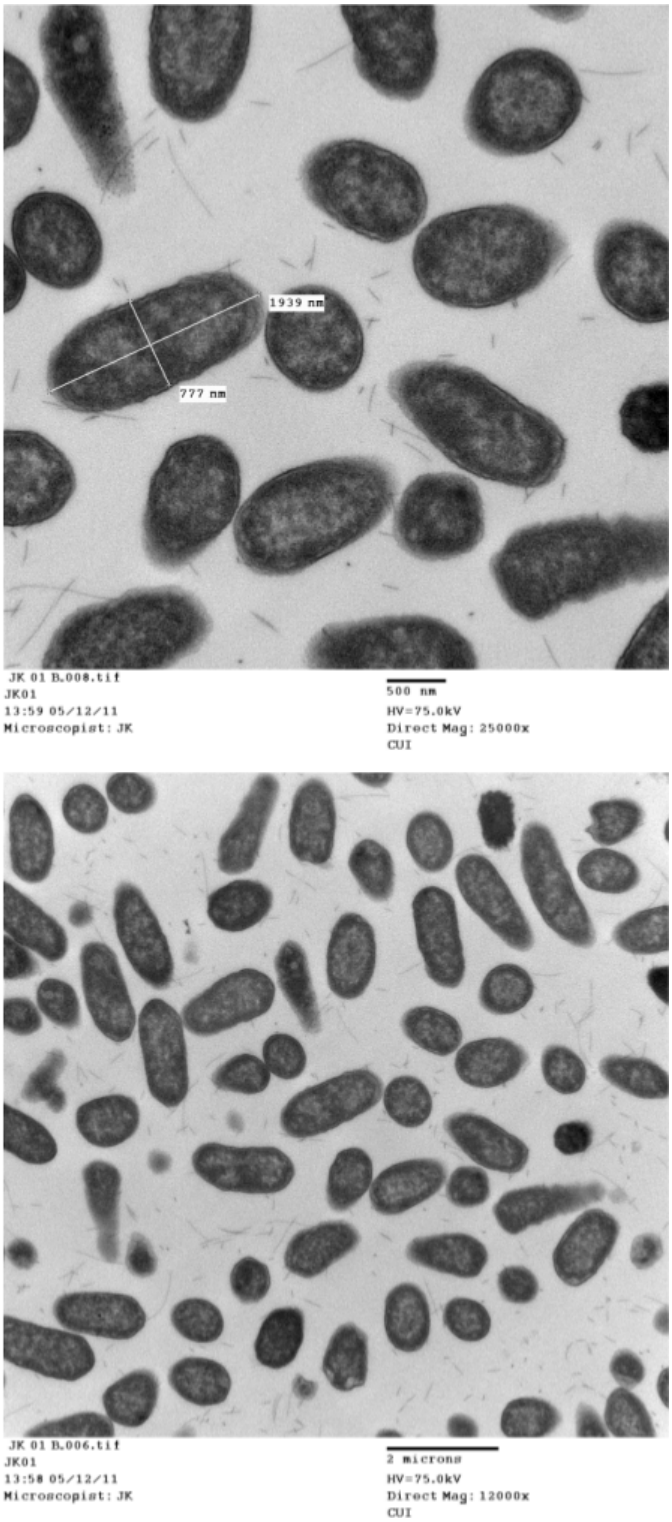


Figure B.1: *TEMs of* E. coli *NCTC 9001 - control cells.*

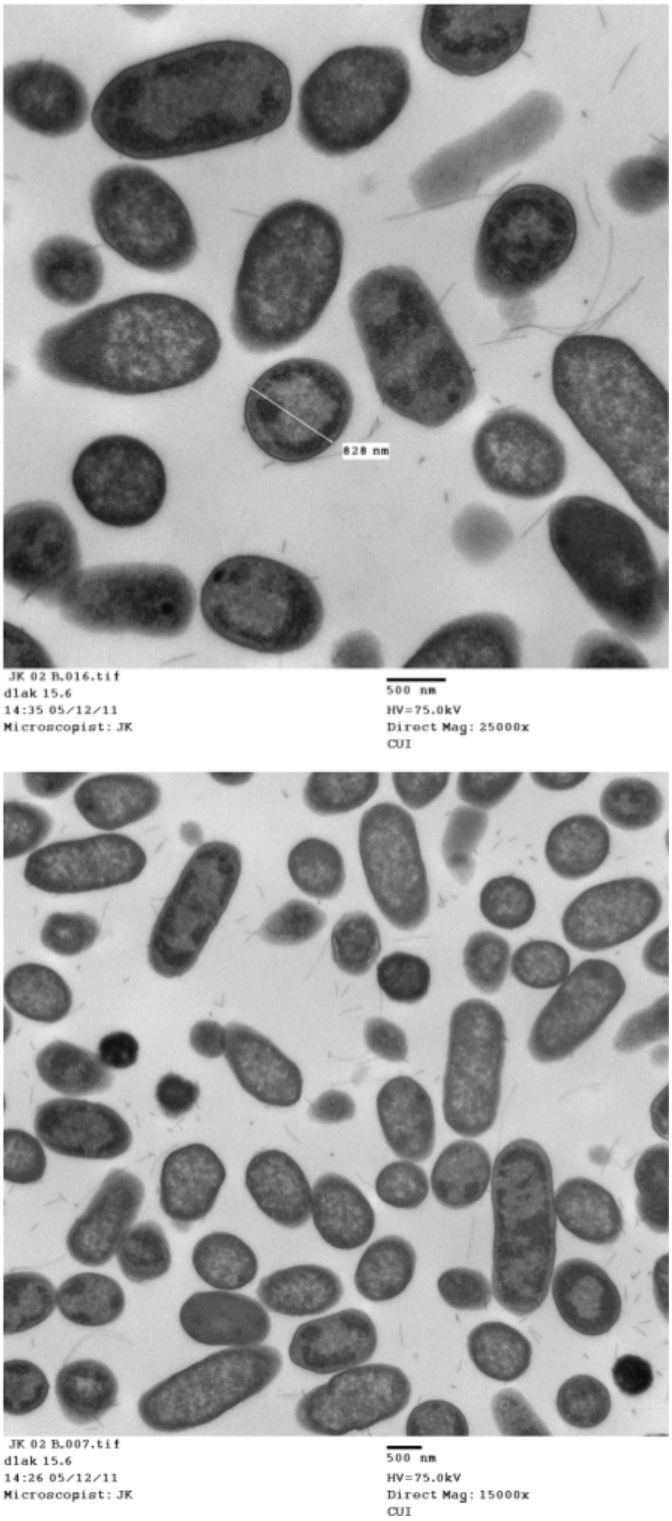


Figure B.2: *TEMs of* E. coli *NCTC 9001 challenged with 15.6 µg/ml D-LAK120-AP13.*

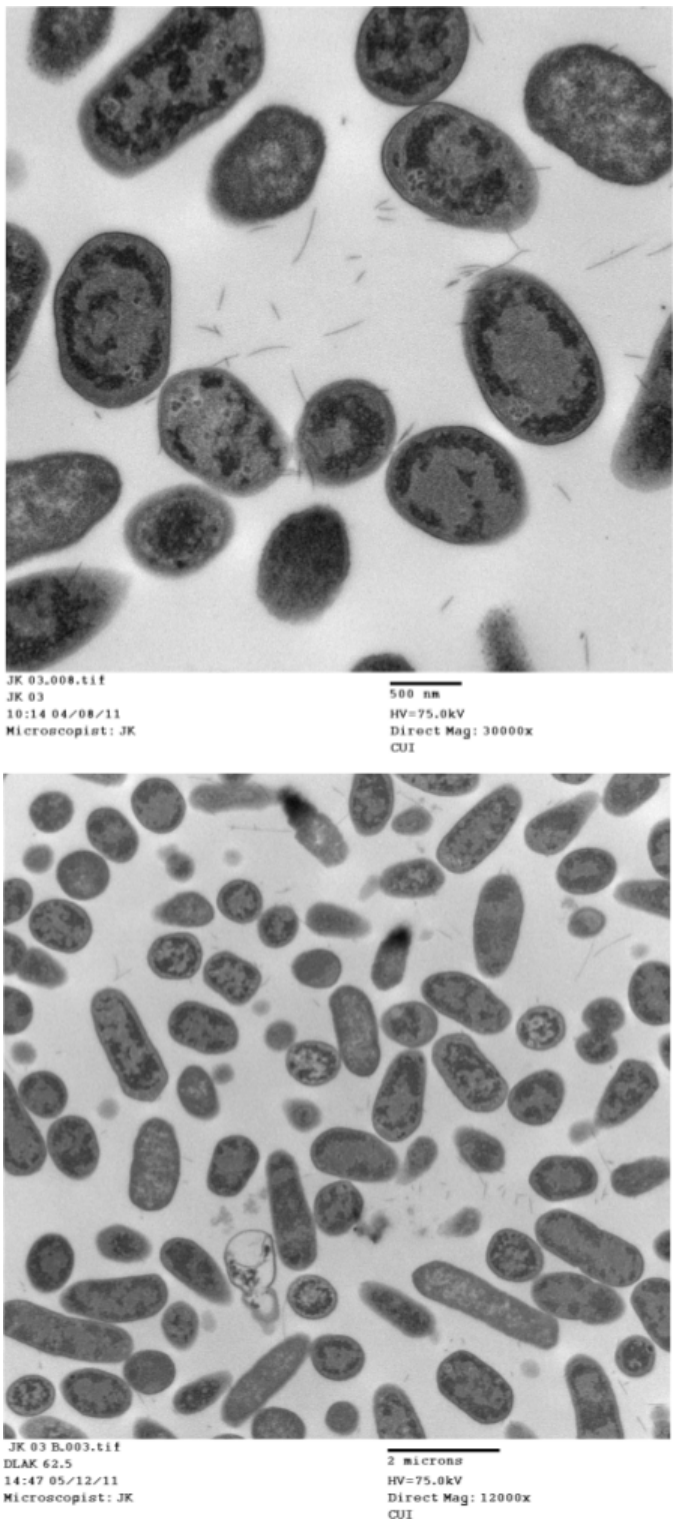


Figure B.3: *TEMs of* E. coli *NCTC 9001 challenged with 62.5 µg/ml D-LAK120-AP13.*

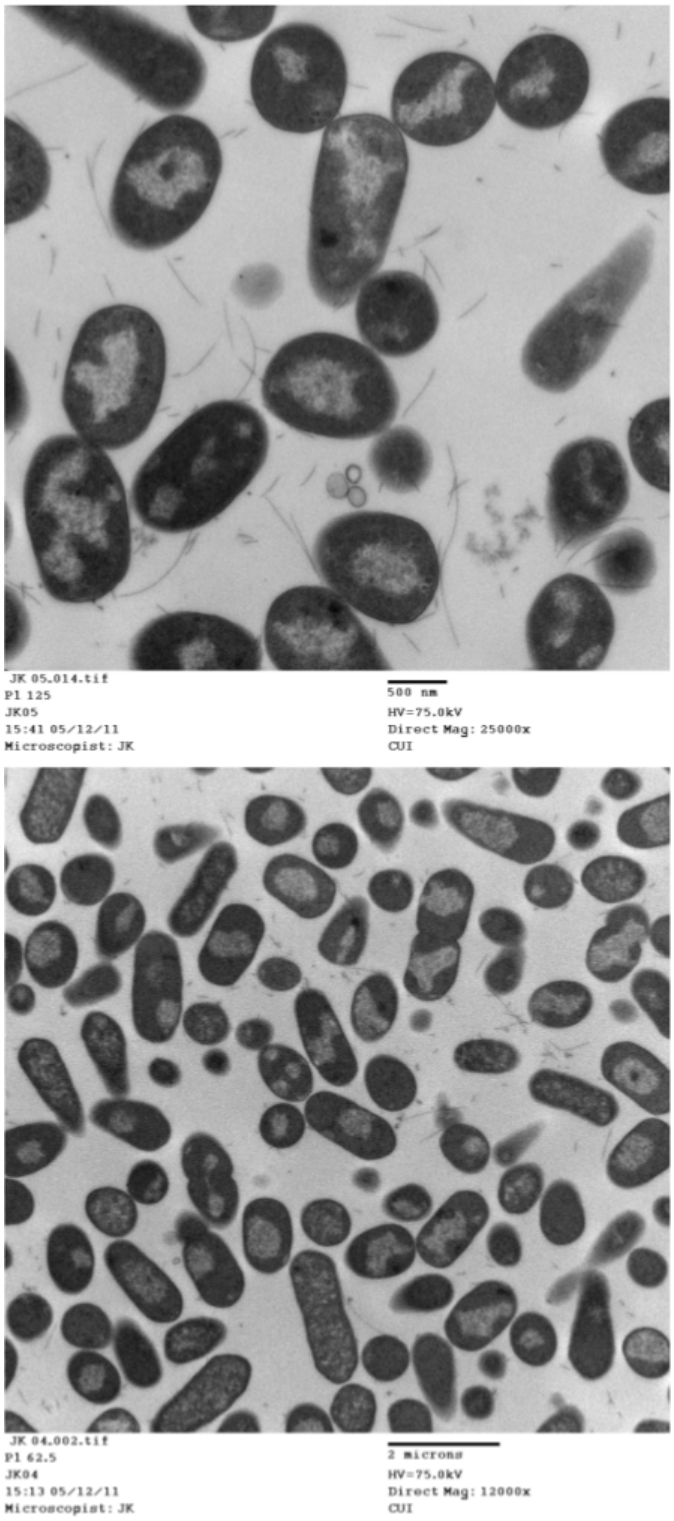


Figure B.4: *TEMs of* E. coli *NCTC 9001 challenged with 62.5 µg/ml pleurocidin.*

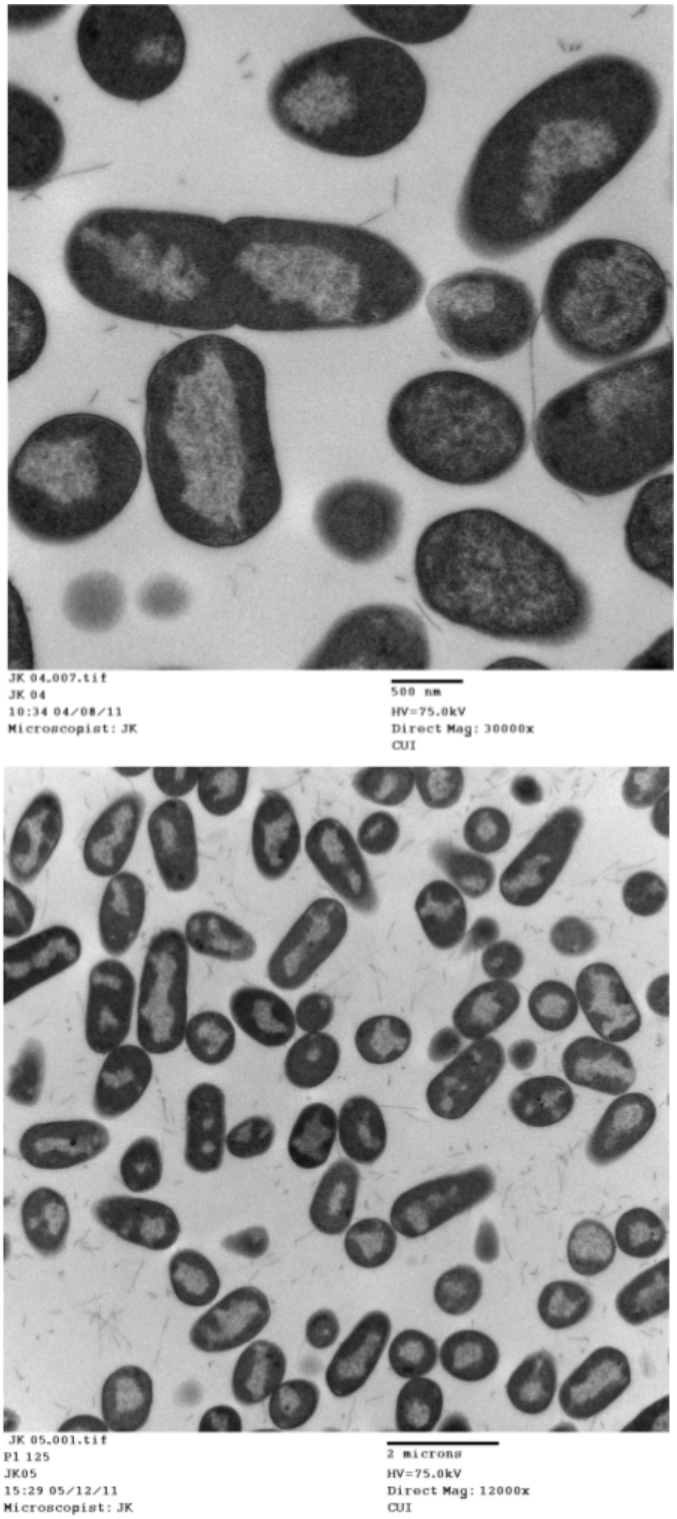


Figure B.5: *TEMs of* E. coli *NCTC 9001 challenged with 125 μg/ml pleurocidin.*

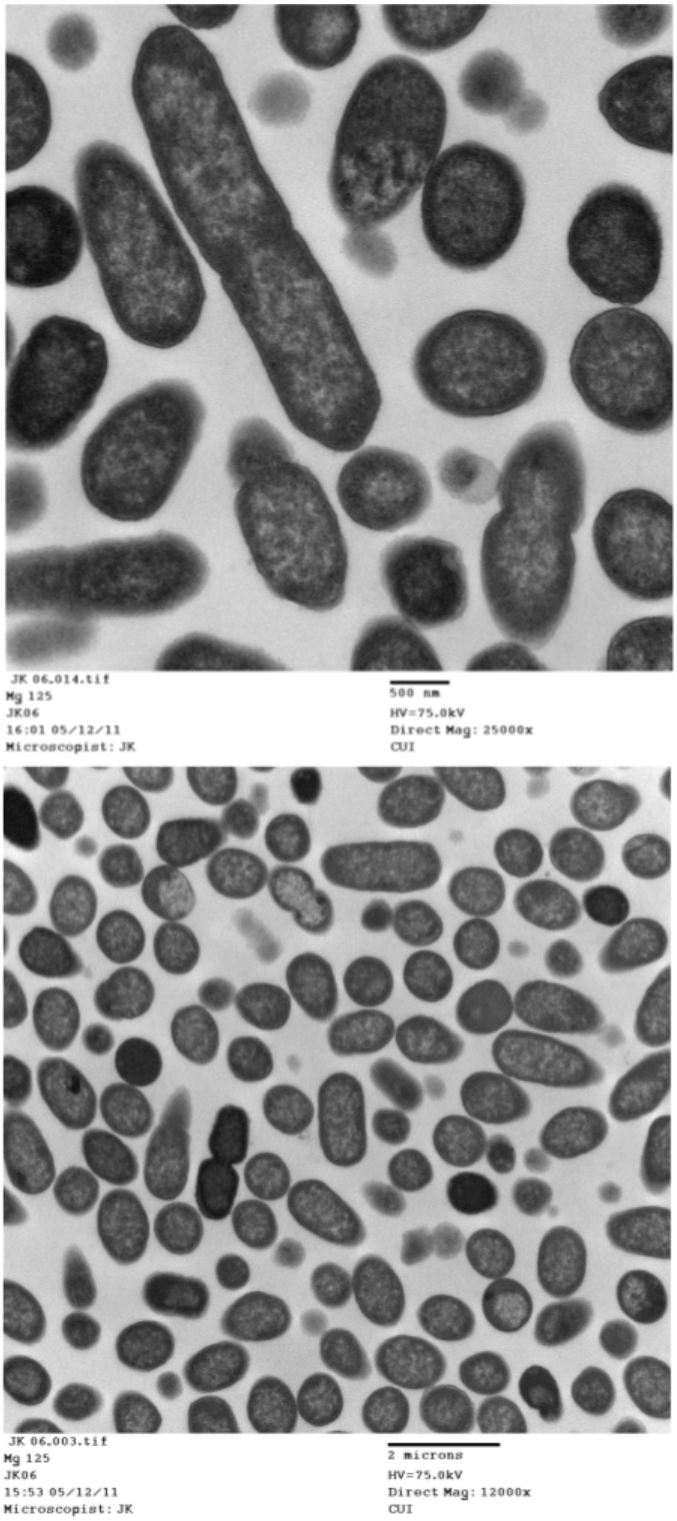


Figure B.6: *TEMs of* E. coli *NCTC 9001 challenged with 125 µg/ml magainin 2.*

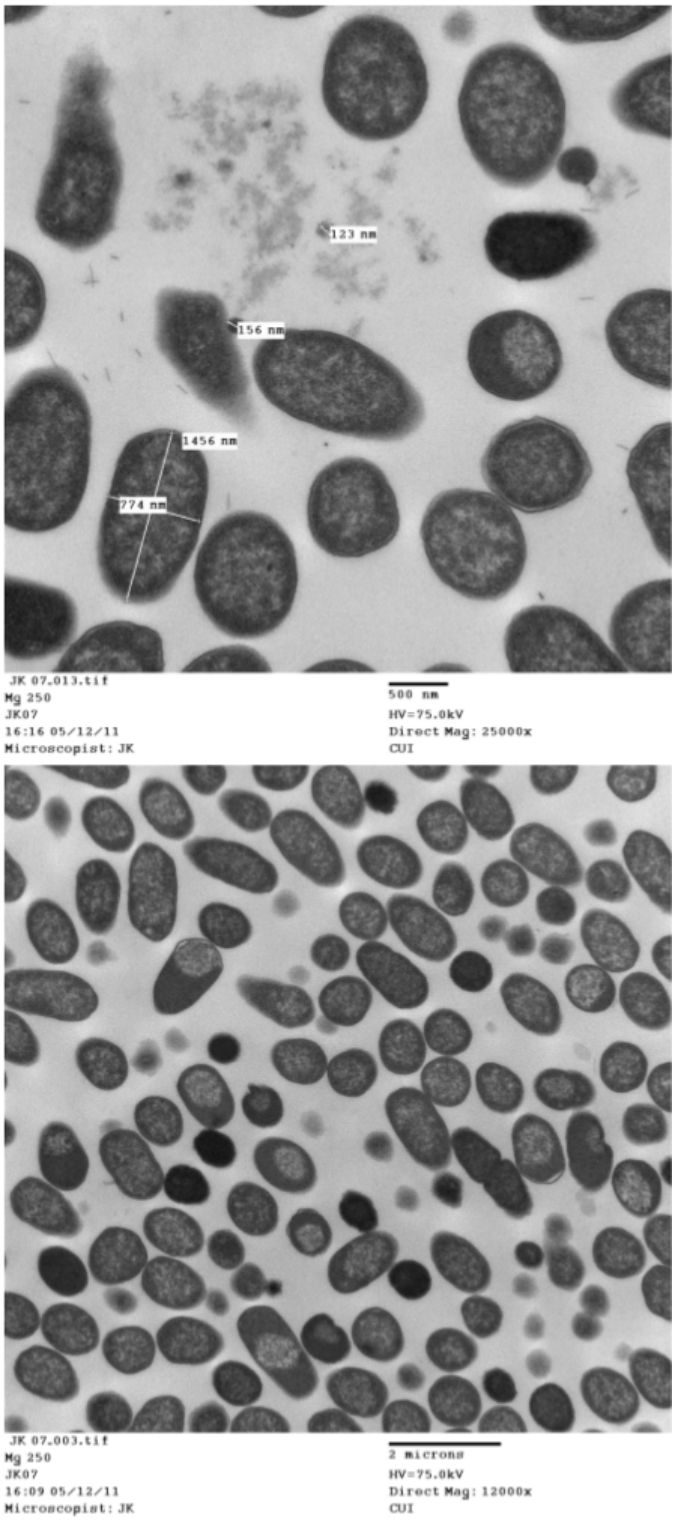


Figure B.7: *TEMs of* E. coli *NCTC 9001 challenged with 250 µg/ml magainin 2.*

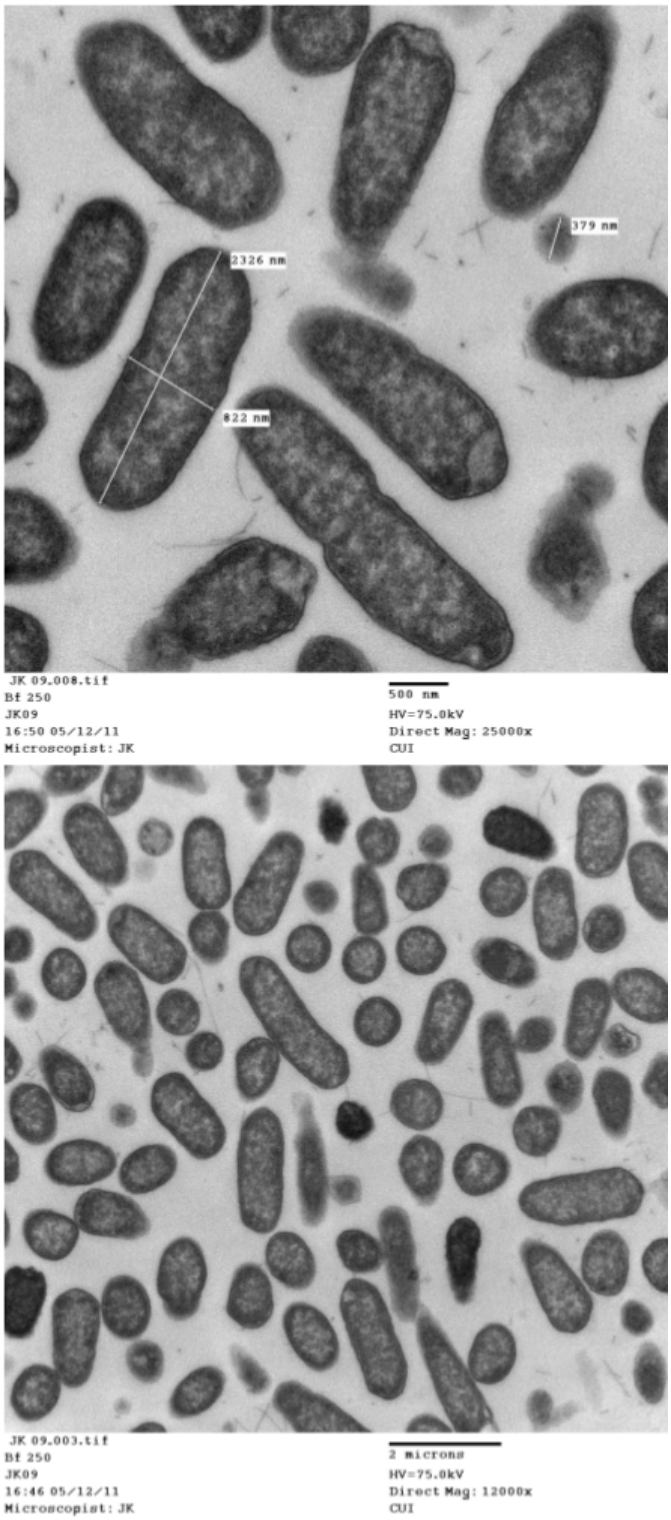


Figure B.8: *TEMs of E. coli NCTC 9001 challenged with 250 µg/ml buforin II.*

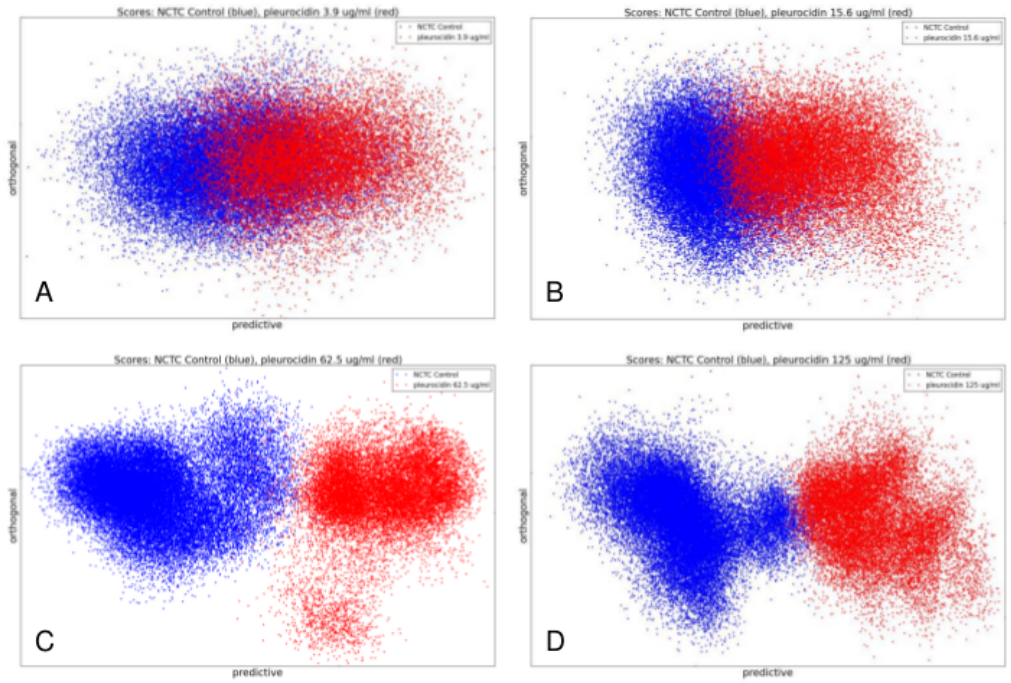


Figure B.9: *OPLS-DA scores for comparisons of* ^{1}H *HR-MAS NMR spectra of control* E. coli *NCTC 9001 and those challenged with pleurocidin at 3.9 µg/ml (A), 15.6 µg/ml (B), 62.5 µg/ml (C) and 125 µg/ml (D). In all panels blue dots represent scores from unchallenged bacteria while red dots represent scores from the respective treatments.*

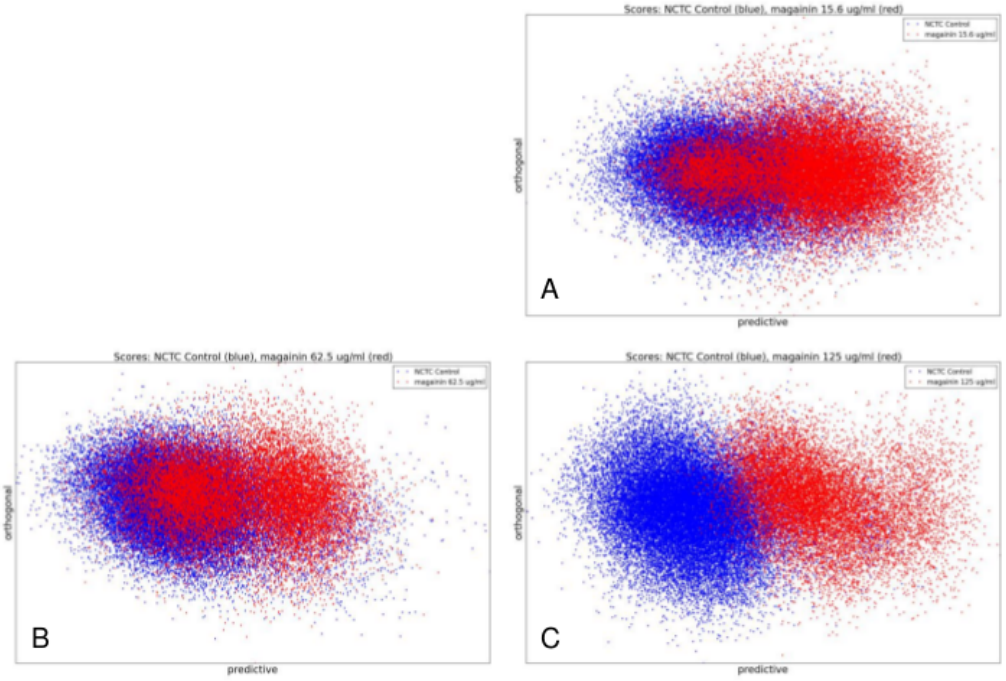


Figure B.10: *OPLS-DA scores for comparisons of 1H HR-MAS NMR spectra of control* E. coli *NCTC 9001 and those challenged with magainin 2 at 15.6 µg/ml (A), 62.5 µg/ml (B) and 125 µg/ml (C). In all panels blue dots represent scores from unchallenged bacteria while red dots represent scores from the respective treatments.*

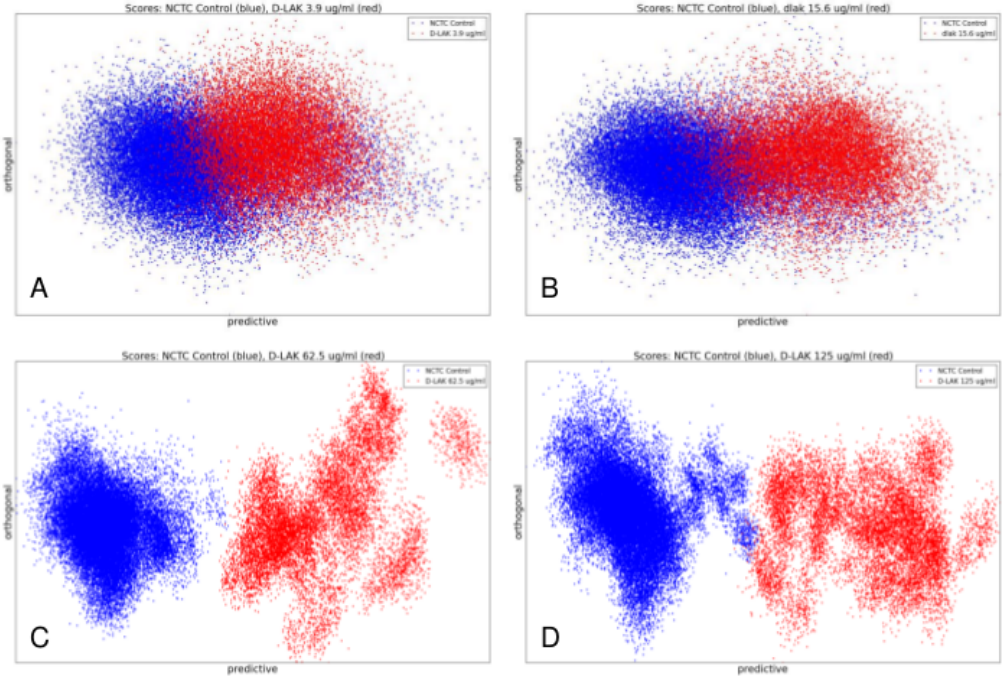


Figure B.11: *OPLS-DA scores for comparisons of* ^{1}*H HR-MAS NMR spectra of control* E. coli *NCTC 9001 and those challenged with D-LAK120-AP13 at 3.9 µg/ml (A), 15.6 µg/ml (B), 62.5 µg/ml (C) and 125 µg/ml (D). In all panels blue dots represent scores from unchallenged bacteria while red dots represent scores from the respective treatments.*

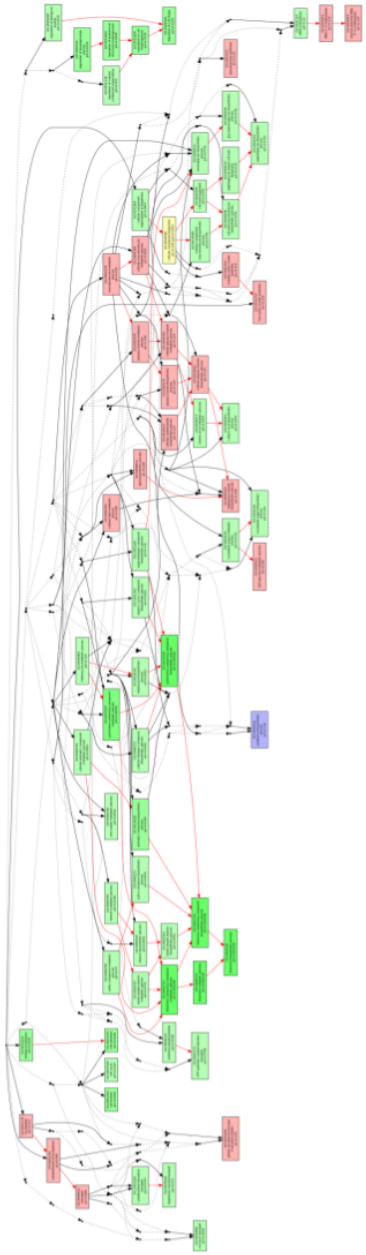


Figure B.12: *Multi GOEAST comparison of biological processes in differentially expressed genes of* E. coli *NCTC 9001 in response to challenge with pleurocidin (red), magainin 2, (blue) and buforin II (green) as detected by the GeneChip*® E. coli *Genome 2.0 Array. Bacteria were challenged for 30 minutes with AMPs at the threshold concentration that elicits a bacterial response as determined by the 1H NMR metabolomic study; 250 µg/ml buforin II, 62.5 µg/ml pleurocidin and 125 µg/ml magainin 2.*

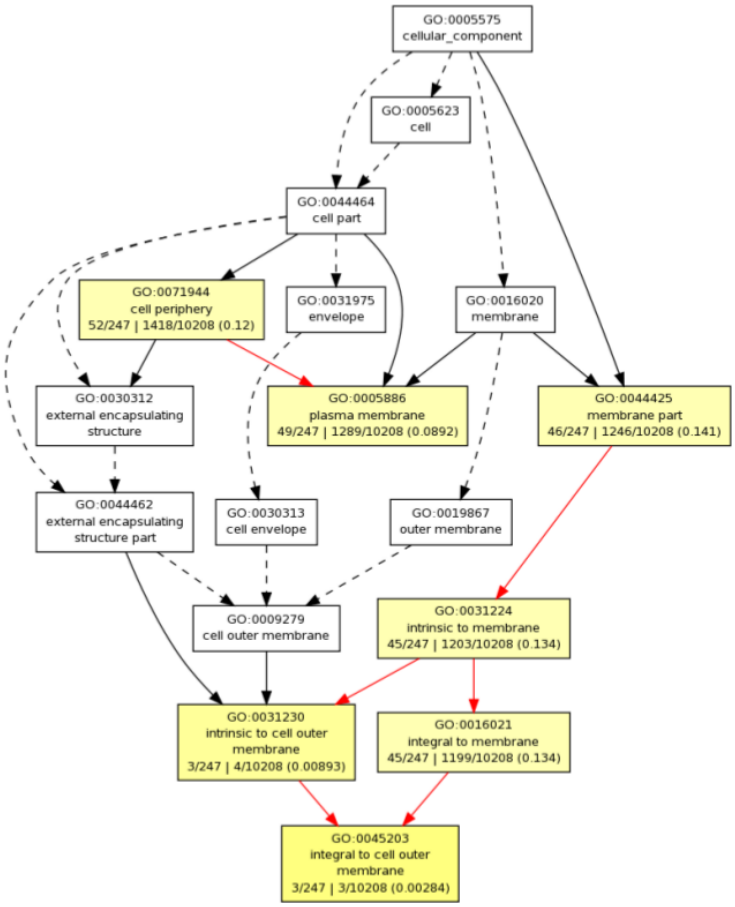


Figure B.13: *GOEAST analysis of cellular component in differentially expressed genes of* E. coli *NCTC 9001 in response to challenge with magainin 2 as detected by the GeneChip*® E. coli *Genome 2.0 Array. Bacteria were challenged with 125 µg/ml magainin 2; the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study.*

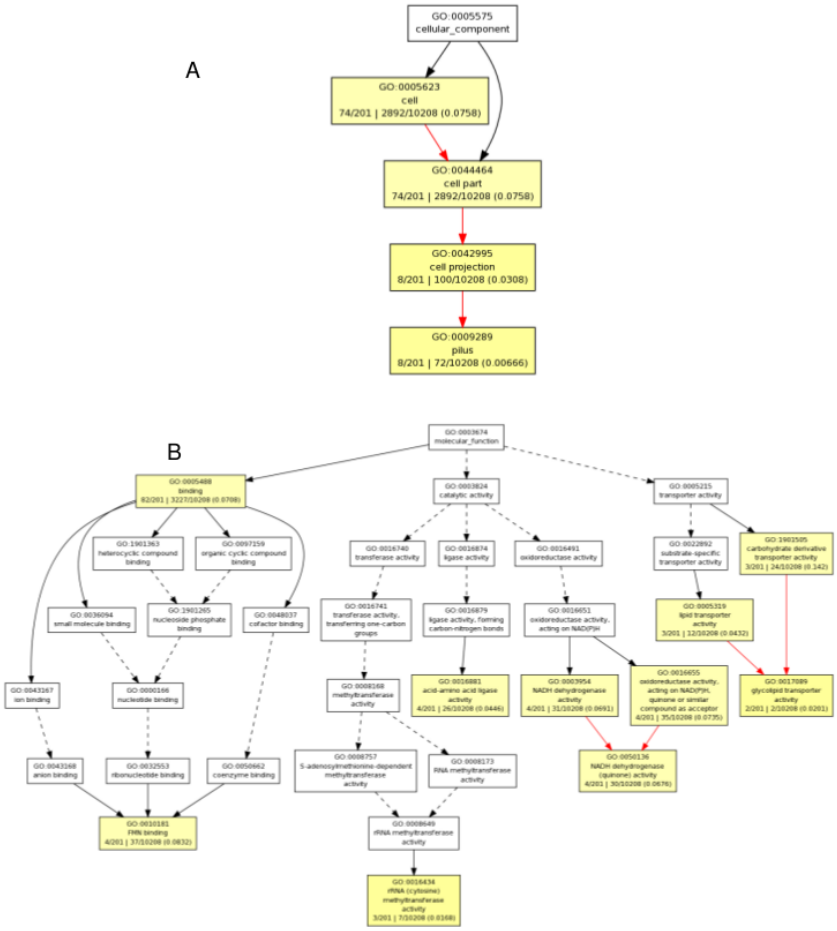


Figure B.14: *GOEAST analysis of cellular component (A) and molecular function (B) in differentially expressed genes of* E. coli *NCTC 9001 in response to challenge with buforin II as detected by the GeneChip® E. coli Genome 2.0 Array. Bacteria were challenged with 250 µg/ml buforin II; the threshold concentration that elicits a bacterial response as determined by the* ^1H *NMR metabolomic study. Note the concentration of genes in cellular component GO terms cell or cell part and in molecular function GO:0005488 binding.*

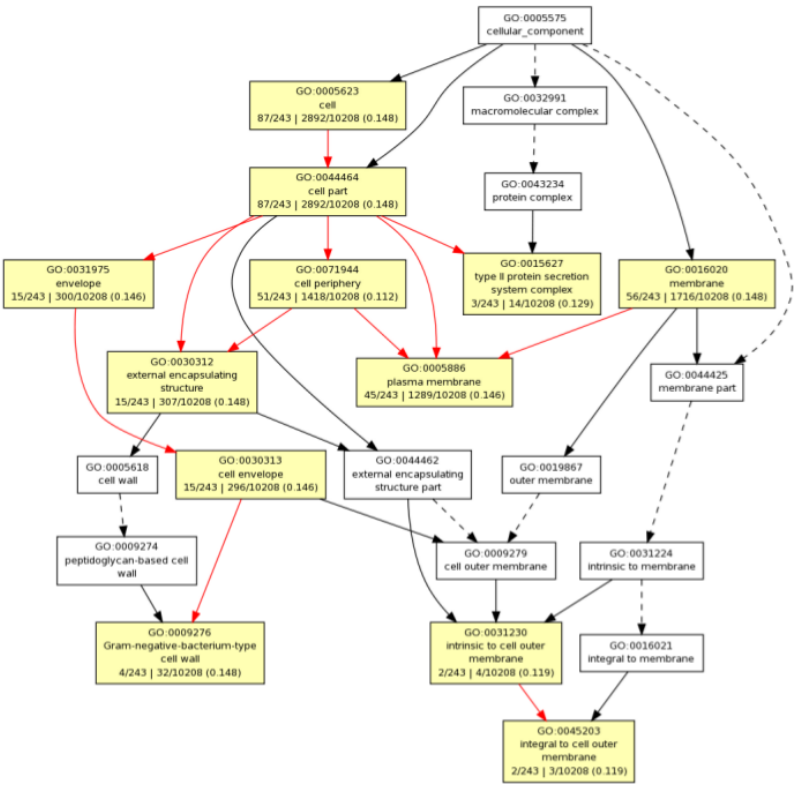


Figure B.15: *GOEAST analysis of cellular component in differentially expressed genes of* E. coli *NCTC 9001 in response to challenge with pleurocidin as detected by the GeneChip*® E. coli *Genome 2.0 Array. Bacteria were challenged with 62.5 µg/ml pleurocidin; the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study. Note the distribution of genes between GO terms cell/cell part, cell periphery and membrane/plasma membrane.*

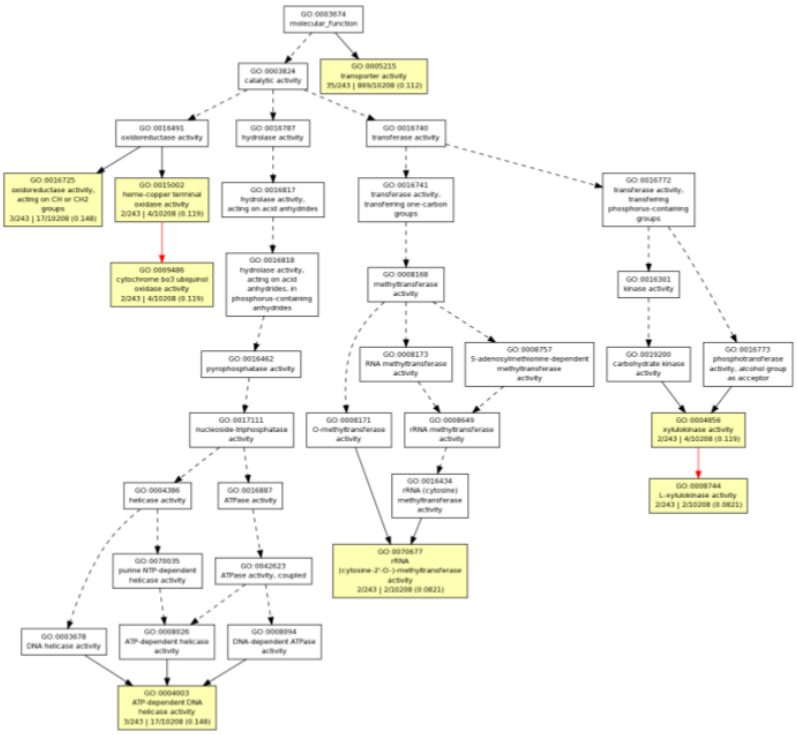


Figure B.16: *GOEAST analysis of molecular function in differentially expressed genes of* E. coli *NCTC 9001 in response to challenge with pleurocidin as detected by the GeneChip*® E. coli *Genome 2.0 Array. Bacteria were challenged with 62.5 µg/ml pleurocidin; the threshold concentration that elicits a bacterial response as determined by the* ^{1}H *NMR metabolomic study. Note the high number of genes corresponding to GO:0005215 transporter activity.*

Bibliography

[1] D. Eisenberg, R. M. Weiss, and T. C. Terwilliger, "The helical hydrophobic moment: a measure of the amphiphilicity of a helix." *Nature*, vol. 299, no. 5881, pp. 371–374, Sep. 1982.

[2] A. P. L. T. Group. (2004) Antimicrobial Sequences Database (ASDb). [Online]. Available: http://www.bbcm.units.it/~tossi/amsdb.html

[3] National institutes of Health, "NIH Human Microbiome Project defines normal bacterial makeup of the body," 2013. [Online]. Available: http://www.nih.gov/news/health/jun2012/nhgri-13.htm

[4] C. V. Forst, "Host-pathogen systems biology." *Drug discovery today*, vol. 11, no. 5-6, pp. 220–7, Mar. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16580599

[5] D. S. Schneider and J. S. Ayres, "Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases." *Nature reviews. Immunology*, vol. 8, no. 11, pp. 889–95, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1038/nri2432

[6] F. Baquero, T. M. Coque, and R. Cantón, "Counteracting antibiotic resistance: breaking barriers among antibacterial strategies," Jul. 2014. [Online]. Available: http://informahealthcare.com/doi/abs/10.1517/14728222.2014.925881

[7] W. B. Dunn and D. I. Ellis, "Metabolomics: Current analytical platforms and methodologies," *TrAC Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, Apr. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165993605000348

[8] V. Dhingra, M. Gupta, T. Andacht, and Z. F. Fu, "New frontiers in proteomics research: a perspective." *International journal of pharmaceutics*, vol. 299, no. 1-2, pp. 1–18, Aug. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378517305002267

[9] S. Oliver, "Systematic functional analysis of the yeast genome," *Trends in Biotechnology*, vol. 16, no. 9, pp. 373–378, Sep. 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167779998012141

[10] I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martínez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muñiz Rascado, Q. Ong, S. Paley, I. Schröder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen, and P. D. Karp, "EcoCyc: fusing model organism databases with systems biology." *Nucleic acids research*, vol. 41, no. Database issue, pp. D605–12, Jan. 2013. [Online]. Available: http://nar.oxfordjournals.org/content/41/D1/D605

[11] O. Fiehn, "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks." *Comp Funct Genomics*, vol. 2, no. 3, pp. 155–168, 2001. [Online]. Available: http://dx.doi.org/10.1002/cfg.82

[12] A. Casadevall and L.-A. Pirofski, "Virulence factors and their mechanisms of action: the view from a damage-response framework." *Journal of water and health*, vol. 7 Suppl 1, pp. S2–S18, Jan. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19717929

[13] G. Storz and R. Hengge, *Bacterial Stress Responses*, E. K. S. N. I. o. C. H. Gisela Storz and F. U. B. B. G. Human Development National Institutes of Health Regine Hengge, Eds. ASM Press, 2010.

[14] J. D. Helmann and M. J. Chamberlin, "Structure and function of bacterial sigma factors." *Annual review of biochemistry*, vol. 57, no. 8, pp. 839–72, Jan. 1988. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3052291

[15] P. H. Bradley, M. J. Brauer, J. D. Rabinowitz, and O. G. Troyanskaya, "Coordinated concentration changes of transcripts and metabolites in Saccharomyces cerevisiae." *PLoS Comput Biol*, vol. 5, no. 1, p. e1000270,

Jan. 2009. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1000270

[16] V. Vinson, "Differential Responses to Change," *Science's STKE*, vol. 2007, no. 384, pp. tw146–tw146, Apr. 2007. [Online]. Available: http://stke.sciencemag.org/content/2007/384/tw146.abstract

[17] K. Nakahigashi, Y. Toya, N. Ishii, T. Soga, M. Hasegawa, H. Watanabe, Y. Takai, M. Honma, H. Mori, and M. Tomita, "Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism." *Molecular systems biology*, vol. 5, no. 1, p. 306, Jan. 2009. [Online]. Available: http://msb.embopress.org/content/5/1/306.abstract

[18] A. N. Brooks, D. J. Reiss, A. Allard, W.-J. Wu, D. M. Salvanha, C. L. Plaisier, S. Chandrasekaran, M. Pan, A. Kaur, and N. S. Baliga, "A system-level model for the microbial regulatory genome." *Molecular systems biology*, vol. 10, no. 7, p. 740, Jan. 2014. [Online]. Available: http://msb.embopress.org/content/10/7/740.abstract

[19] F. Wessely, M. Bartl, R. Guthke, P. Li, S. Schuster, and C. Kaleta, "Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs." *Molecular systems biology*, vol. 7, no. 1, p. 515, Jan. 2011. [Online]. Available: http://msb.embopress.org/content/7/1/515.abstract

[20] A. Casadevall and L. Pirofski, "Host-pathogen interactions: the attributes of virulence." *The Journal of infectious diseases*, vol. 184, no. 3, pp. 337–44, Aug. 2001. [Online]. Available: http://jid.oxfordjournals.org/content/184/3/337.long

[21] J. D. Bryers, "Medical biofilms." *Biotechnology and bioengineering*, vol. 100, no. 1, pp. 1–18, May 2008. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2706312&tool=pmcentrez&rendertype=abstract

[22] P. S. Stewart and M. J. Franklin, "Physiological heterogeneity in biofilms." *Nature reviews. Microbiology*, vol. 6, no. 3, pp. 199–210, Mar. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18264116

[23] J. W. Costerton, "Bacterial Biofilms: A Common Cause of Persistent Infections," *Science*, vol. 284, no. 5418, pp. 1318–1322, May 1999. [Online]. Available: http://www.sciencemag.org/content/284/5418/1318

[24] J. Kozlowska, D. W. Rivett, L. S. Vermeer, M. P. Carroll, K. D. Bruce, A. J. Mason, and G. B. Rogers, "A relationship between Pseudomonal growth behaviour and cystic fibrosis patient lung function identified in a metabolomic investigation." *Metabolomics : Official journal of the Metabolomic Society*, vol. 9, no. 6, Dec. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24367285

[25] H. v. Jenssen, P. Hamill, and R. E. W. Hancock, "Peptide antimicrobial agents." *Clin Microbiol Rev*, vol. 19, no. 3, pp. 491–511, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1128/CMR.00056-05

[26] R. E. Ley, D. A. Peterson, and J. I. Gordon, "Ecological and evolutionary forces shaping microbial diversity in the human intestine." *Cell*, vol. 124, no. 4, pp. 837–48, Feb. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092867406001929

[27] J. Lederberg, "Infectious History," *Science*, vol. 288, no. 5464, pp. 287–293, Apr. 2000. [Online]. Available: http://www.sciencemag.org/content/288/5464/287.full

[28] F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon, "Host-bacterial mutualism in the human intestine." *Science (New York, N.Y.)*, vol. 307, no. 5717, pp. 1915–20, Mar. 2005. [Online]. Available: http://www.sciencemag.org/content/307/5717/1915

[29] V. K. Ridaura, J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, A. L. Kau, N. W. Griffin, V. Lombard, B. Henrissat, J. R. Bain, M. J. Muehlbauer, O. Ilkayeva, C. F. Semenkovich, K. Funai, D. K. Hayashi, B. J. Lyle, M. C. Martini, L. K. Ursell, J. C. Clemente, W. Van Treuren, W. A. Walters, R. Knight, C. B. Newgard, A. C. Heath, and J. I. Gordon, "Gut microbiota from twins discordant for obesity modulate metabolism in mice." *Science (New York, N.Y.)*, vol. 341, no. 6150, p. 1241214, Sep. 2013. [Online]. Available: http://www.sciencemag.org/content/341/6150/1241214

[30] L. Cox, S. Yamanishi, J. Sohn, A. Alekseyenko, J. Leung, I. Cho, S. Kim, H. Li, Z. Gao, D. Mahana, J. ZárateRodriguez, A. Rogers, N. Robine, P. Loke, and M. Blaser, "Altering the Intestinal Microbiota during a

Critical Developmental Window Has Lasting Metabolic Consequences," *Cell*, vol. 158, no. 4, pp. 705–721, Aug. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092867414008216

[31] C. E. West, M. C. Jenmalm, and S. L. Prescott, "The gut microbiota and its role in the development of allergic disease: a wider perspective." *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, Apr. 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24773202

[32] G. B. Huffnagle, "The microbiota and allergies/asthma." *PLoS pathogens*, vol. 6, no. 5, p. e1000549, May 2010. [Online]. Available: http://dx.plos.org/10.1371/journal.ppat.1000549

[33] World Allergy Organisation, "White Book on Allergy," Tech. Rep., 2013.

[34] J. Gallagher, "Gut bugs 'help prevent allergies'," 2014. [Online]. Available: http://www.bbc.co.uk/news/health-28887088

[35] P. L. Foster, "Stress-Induced Mutagenesis in Bacteria," Oct. 2008. [Online]. Available: http://informahealthcare.com/doi/full/10.1080/10409230701648494

[36] J. Cairns and P. L. Foster, "Adaptive reversion of a frameshift mutation in Escherichia coli." *Genetics*, vol. 128, no. 4, pp. 695–701, Aug. 1991. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1204544&tool=pmcentrez&rendertype=abstract

[37] R. Lange and R. Hengge-Aronis, "Growth phase-regulated expression of bolA and morphology of stationary-phase Escherichia coli cells are controlled by the novel sigma factor sigma S." *Journal of bacteriology*, vol. 173, no. 14, pp. 4474–81, Jul. 1991. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=208111&tool=pmcentrez&rendertype=abstract

[38] G. Schreiber, E. Z. Ron, and G. Glaser, "ppGpp-mediated regulation of DNA replication and cell division in Escherichia coli," *Current Microbiology*, vol. 30, no. 1, pp. 27–32, Jan. 1995. [Online]. Available: http://link.springer.com/10.1007/BF00294520

[39] S. Carneiro, A. Lourenço, E. C. Ferreira, and I. Rocha, "Stringent response of Escherichia coli: revisiting the bibliome using literature mining."

Microbial informatics and experimentation, vol. 1, no. 1, p. 14, Jan. 2011. [Online]. Available: http://www.microbialinformaticsj.com/content/1/1/14

[40] E. K. O'Reilly and K. N. Kreuzer, "Isolation of SOS constitutive mutants of Escherichia coli." *Journal of bacteriology*, vol. 186, no. 21, pp. 7149–60, Nov. 2004. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=523190&tool=pmcentrez&rendertype=abstract

[41] J. Hong, J.-M. Ahn, B. C. Kim, and M. B. Gu, "Construction of a functional network for common DNA damage responses in Escherichia coli." *Genomics*, vol. 93, no. 6, pp. 514–24, Jun. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19232389

[42] R. Sikdar, A. R. Simmons, and W. T. Doerrler, "Multiple envelope stress response pathways are activated in an Escherichia coli strain with mutations in two members of the DedA membrane protein family." *Journal of bacteriology*, vol. 195, no. 1, pp. 12–24, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3536178&tool=pmcentrez&rendertype=abstract

[43] S. Bury-Moné, Y. Nomane, N. Reymond, R. Barbet, E. Jacquet, S. Imbeaud, A. Jacq, and P. Bouloc, "Global analysis of extracytoplasmic stress signaling in Escherichia coli." *PLoS genetics*, vol. 5, no. 9, p. e1000651, Sep. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2731931&tool=pmcentrez&rendertype=abstract

[44] J. A. Imlay, "Cellular defenses against superoxide and hydrogen peroxide." *Annual review of biochemistry*, vol. 77, pp. 755–76, Jan. 2008. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3057177&tool=pmcentrez&rendertype=abstract

[45] G. Storz and J. A. Imlayt, "Oxidative stress," *Current Opinion in Microbiology*, vol. 2, no. 2, pp. 188–194, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1369527499800332

[46] S. Fraud and K. Poole, "Oxidative stress induction of the MexXY multidrug efflux genes and promotion of aminoglycoside resistance development in Pseudomonas aeruginosa." *Antimicrobial agents and chemotherapy*, vol. 55, no. 3, pp. 1068–74, Mar. 2011. [Online]. Available: http://aac.asm.org/content/55/3/1068

[47] J. A. Imlay, S. M. Chin, and S. Linn, "Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro." *Science (New York, N.Y.)*, vol. 240, no. 4852, pp. 640–2, Apr. 1988. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2834821

[48] C. E. Donnelly and G. C. Walker, "groE mutants of Escherichia coli are defective in umuDC-dependent UV mutagenesis." *J. Bacteriol.*, vol. 171, no. 11, pp. 6117–6125, Nov. 1989. [Online]. Available: http://jb.asm.org/content/171/11/6117

[49] J. C. Layton and P. L. Foster, "Error-prone DNA polymerase IV is regulated by the heat shock chaperone GroE in Escherichia coli." *Journal of bacteriology*, vol. 187, no. 2, pp. 449–57, Jan. 2005. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=543561&tool=pmcentrez&rendertype=abstract

[50] L. N. Csonka, "Physiological and genetic responses of bacteria to osmotic stress." *Microbiol. Mol. Biol. Rev.*, vol. 53, no. 1, pp. 121–147, Mar. 1989. [Online]. Available: http://mmbr.asm.org/content/53/1/121.abstract

[51] O. Fiehn, "Metabolomics the link between genotypes and phenotypes," *Plant Molecular Biology*, vol. 48, no. 1-2, pp. 155–171, Jan. 2002. [Online]. Available: http://link.springer.com/article/10.1023/A%3A1013713905833

[52] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data." *Xenobiotica; the fate of foreign compounds in biological systems*, vol. 29, no. 11, pp. 1181–9, Nov. 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10598751

[53] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin, "Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy." *Chemical Society reviews*, vol. 40, no. 1, pp. 387–426, Jan. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20717559

[54] B. H. ter Kuile and H. V. Westerhoff, "Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway." *FEBS letters*, vol. 500, no. 3, pp. 169–71, Jul. 2001. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11445079

[55] J. L. Griffin and J. P. Shockcor, "Metabolic profiles of cancer cells." *Nature reviews. Cancer*, vol. 4, no. 7, pp. 551–61, Jul. 2004. [Online]. Available: http://dx.doi.org/10.1038/nrc1390

[56] D. G. Robertson, M. D. Reily, and J. D. Baker, "Metabonomics in pharmaceutical discovery and development." *J Proteome Res*, vol. 6, no. 2, pp. 526–539, Feb. 2007. [Online]. Available: http://dx.doi.org/10.1021/pr060535c

[57] B. Lahner, J. Gong, M. Mahmoudian, E. L. Smith, K. B. Abid, E. E. Rogers, M. L. Guerinot, J. F. Harper, J. M. Ward, L. McIntyre, J. I. Schroeder, and D. E. Salt, "Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana." *Nature biotechnology*, vol. 21, no. 10, pp. 1215–21, Oct. 2003. [Online]. Available: http://dx.doi.org/10.1038/nbt865

[58] A. C. Guo, T. Jewison, M. Wilson, Y. Liu, C. Knox, Y. Djoumbou, P. Lo, R. Mandal, R. Krishnamurthy, and D. S. Wishart, "ECMDB: The E. coli Metabolome Database," *Nucleic Acids Research*, Oct. 2012. [Online]. Available: http://nar.oxfordjournals.org/content/early/2012/10/28/nar.gks992.full%#ref-23

[59] J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, and D. B. Kell, "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting." *Nature biotechnology*, vol. 21, no. 6, pp. 692–6, Jun. 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12740584

[60] D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic, and S. G. Oliver, "Metabolic footprinting and systems biology: the medium is the message." *Nat Rev Microbiol*, vol. 3, no. 7, pp. 557–565, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1038/nrmicro1177

[61] N. D. Jayavelu and N. S. Bar, "Metabolomic studies of human gastric cancer: Review." *World journal of gastroenterology : WJG*, vol. 20, no. 25, pp. 8092–8101, Jul. 2014. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4081680&tool=pmcentrez&rendertype=abstract

[62] J. P. Grivet, A. M. Delort, and J. C. Portais, "NMR and microbiology: from physiology to metabolomics." *Biochimie*, vol. 85, no. 9, pp. 823–840, Sep. 2003.

[63] M. Baker, "Metabolomics: from small molecules to big ideas," *Nature Methods*, vol. 8, no. 2, pp. 117–121, Feb. 2011. [Online]. Available: http://dx.doi.org/10.1038/nmeth0211-117

[64] Y. Lan, Y. Ye, J. Kozlowska, J. K. W. Lam, A. F. Drake, and A. J. Mason, "Structural contributions to the intracellular targeting strategies of antimicrobial peptides." *Biochimica et biophysica acta*, vol. 1798, no. 10, pp. 1934–43, Oct. 2010. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3309561&tool=pmcentrez&rendertype=abstract

[65] R. Kramer, *Chemometric Techniques for Quantitative Analysis(Google eBook)*. CRC Press, 1998. [Online]. Available: http://books.google.com/books?id=iBpOzwAOfHYC&pgis=1

[66] G. Nyamundanda, I. C. Gormley, Y. Fan, W. M. Gallagher, and L. Brennan, "MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach." *BMC bioinformatics*, vol. 14, no. 1, p. 338, Jan. 2013. [Online]. Available: http://www.biomedcentral.com/1471-2105/14/338

[67] R. T. Eakin, L. O. Morgan, C. T. Gregg, and N. A. Matwiyoff, "Carbon-13 nuclear magnetic resonance spectroscopy of living cells and their metabolism of a specifically labeled 13C substrate." *FEBS Lett*, vol. 28, no. 3, pp. 259–264, Dec. 1972.

[68] J. C. Lindon, E. Holmes, and J. K. Nicholson, "Peer Reviewed: So Whats the Deal with Metabonomics?" *Analytical Chemistry*, vol. 75, no. 17, pp. 384 A–391 A, Sep. 2003. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ac031386%2B

[69] F. Bloch, W. Hansen, and M. Packard, "Nuclear Induction," *Physical Review*, vol. 69, no. 3-4, pp. 127–127, Feb. 1946. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRev.69.127

[70] E. Purcell, H. Torrey, and R. Pound, "Resonance Absorption by Nuclear Magnetic Moments in a Solid," *Physical Review*, vol. 69, no. 1-2, pp. 37–38, Jan. 1946. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRev.69.37

[71] J. Emsley and J. Feeney, "Forty years of Progress in Nuclear Magnetic Resonance Spectroscopy," *Progress in Nuclear Magnetic Resonance*

Spectroscopy, vol. 50, no. 4, pp. 179–198, Jul. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0079656507000039

[72] M. J. Duer, *Introduction to Solid-State NMR Spectroscopy: Amazon.co.uk: : Books*, 2nd ed., 2008. [Online]. Available: http://www.amazon.co.uk/Introduction-Solid-State-Spectroscopy-Melinda-Duer/dp/1405109149

[73] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, 2nd ed., ser. Tetrahedron Organic Chemistry Series. Elsevier, 2009, vol. 27. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1460156708100034

[74] H. Kovacs, D. Moskau, and M. Spraul, "Cryogenically cooled probesa leap in NMR technology," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 46, no. 2-3, pp. 131–155, May 2005. [Online]. Available: http://www.journals.elsevierhealth.com/periodicals/jpnmrs/article/S0079-6565(05)00017-8/fulltext

[75] T. Hwang and A. Shaka, "Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients," *Journal of Magnetic Resonance, Series A*, vol. 112, no. 2, pp. 275–279, Feb. 1995. [Online]. Available: http://www.readcube.com/articles/10.1006/jmra.1995.1047?locale=en

[76] M. Piotto, V. Saudek, and V. Sklená, "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions," *Journal of Biomolecular NMR*, vol. 2, no. 6, pp. 661–665, Nov. 1992. [Online]. Available: http://link.springer.com/10.1007/BF02192855

[77] R. T. Mckay, "How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution," *Concepts in Magnetic Resonance Part A*, vol. 38A, no. 5, pp. 197–220, Sep. 2011. [Online]. Available: http://doi.wiley.com/10.1002/cmr.a.20223

[78] S. Meiboom and D. Gill, "Modified Spin-Echo Method for Measuring Nuclear Relaxation Times," *Review of Scientific Instruments*, vol. 29, no. 8, p. 688, 1958. [Online]. Available: http://adsabs.harvard.edu/abs/1958RScI...29..688M

[79] M. R. Viant, "Improved methods for the acquisition and interpretation of NMR metabolomic data." *Biochem Biophys Res Commun*, vol. 310, no. 3, pp. 943–948, Oct. 2003.

[80] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2008. [Online]. Available: http://www.amazon.co.uk/Spin-Dynamics-Nuclear-Magnetic-Resonance/dp/0470511176

[81] J. Keeler, *Understanding NMR Spectroscopy*, 2010. [Online]. Available: http://www.amazon.co.uk/Understanding-NMR-Spectroscopy-James-Keeler/dp/0470746084

[82] O. Beckonert, M. Coen, H. C. Keun, Y. Wang, T. M. D. Ebbels, E. Holmes, J. C. Lindon, and J. K. Nicholson, "High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues." *Nature protocols*, vol. 5, no. 6, pp. 1019–32, Jun. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20539278

[83] M. Renault, L. Shintu, M. Piotto, and S. Caldarelli, "Slow-spinning low-sideband HR-MAS NMR spectroscopy: delicate analysis of biological samples." *Scientific reports*, vol. 3, p. 3349, Jan. 2013. [Online]. Available: http://www.nature.com/srep/2013/131127/srep03349/full/srep03349.html

[84] P. Tripathi, B. S. Somashekar, M. Ponnusamy, A. Gursky, S. Dailey, P. Kunju, C. T. Lee, A. M. Chinnaiyan, T. M. Rajendiran, and A. Ramamoorthy, "HR-MAS NMR tissue metabolomic signatures cross-validated by mass spectrometry distinguish bladder cancer from benign disease." *Journal of proteome research*, vol. 12, no. 7, pp. 3519–28, Jul. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3722911&tool=pmcentrez&rendertype=abstract

[85] R. Mirnezami, B. Jiménez, J. V. Li, J. M. Kinross, K. Veselkov, R. D. Goldin, E. Holmes, J. K. Nicholson, and A. Darzi, "Rapid diagnosis and staging of colorectal cancer via high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy of intact tissue biopsies." *Annals of surgery*, vol. 259, no. 6, pp. 1138–49, Jun. 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23860197

[86] M. Chauton, O. Optun, T. Bathen, Z. Volent, I. Gribbestad, and G. Johnsen, "HR MAS 1H NMR spectroscopy analysis of marine microalgal

whole cells," *Marine Ecology Progress Series*, vol. 256, pp. 57–62, Jul. 2003. [Online]. Available: http://brage.bibsys.no/xmlui/handle/11250/108784

[87] H. Hanana, G. Simon, N. Kervarec, B. A. Mohammadou, and S. Cérantola, "HRMAS NMR as a tool to study metabolic responses in heart clam Ruditapes decussatus exposed to Roundup." *Talanta*, vol. 97, pp. 425–31, Aug. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22841103

[88] E. R. Andrew, A. Bradbury, and R. G. Eades, "Removal of Dipolar Broadening of Nuclear Magnetic Resonance Spectra of Solids by Specimen Rotation," *Nature*, vol. 183, no. 4678, pp. 1802–1803, Jun. 1959. [Online]. Available: http://dx.doi.org/10.1038/1831802a0

[89] I. Lowe, "Free Induction Decays of Rotating Solids," *Physical Review Letters*, vol. 2, no. 7, pp. 285–287, Apr. 1959. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.2.285

[90] C. Ludwig and M. R. Viant, "Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox." *Phytochemical analysis : PCA*, vol. 21, no. 1, pp. 22–32, Jan. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19904730

[91] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. D. Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel, and I. Forsythe, "HMDB: a knowledgebase for the human metabolome." *Nucleic Acids Res*, vol. 37, no. Database issue, pp. D603—-D610, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1093/nar/gkn810

[92] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, and J. L. Markley, "BioMagResBank," *Nucleic Acids Research*, vol. 36, no. Database issue, pp. D402—-D408, Jan. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2238925/http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2238925/pdf/gkm957.pdf

[93] T. D. C. Laurence M. Harwood, *Introduction to Organic Spectroscopy Oxford Chemistry Primers: Amazon.co.uk: Laurence M. Harwood, Timothy D.W. Claridge: Books*. [Online]. Available: http://www.amazon.co.uk/Introduction-Organic-Spectroscopy-Chemistry-Primers/dp/0198557558

[94] O. Cloarec, M.-E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes, and J. Nicholson, "Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets." *Anal Chem*, vol. 77, no. 5, pp. 1282–1289, Mar. 2005. [Online]. Available: http://dx.doi.org/10.1021/ac048630x

[95] C. Ludwig and U. L. Günther, "MetaboLab–advanced NMR data processing and analysis for metabolomics." *BMC bioinformatics*, vol. 12, no. 1, p. 366, Jan. 2011. [Online]. Available: http://www.biomedcentral.com/1471-2105/12/366

[96] D. Tulpan, S. Léger, L. Belliveau, A. Culf, and M. Cuperlović-Culf, "MetaboHunter: an automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures." *BMC bioinformatics*, vol. 12, no. 1, p. 400, Jan. 2011. [Online]. Available: http://www.biomedcentral.com/1471-2105/12/400

[97] S. Ma and J. Huang, "Regularized gene selection in cancer microarray meta-analysis." *BMC bioinformatics*, vol. 10, p. 1, Jan. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2631520&tool=pmcentrez&rendertype=abstract

[98] A. Alonso, M. A. Rodríguez, M. Vinaixa, R. Tortosa, X. Correig, A. Julià, and S. Marsal, "Focus: a robust workflow for one-dimensional NMR spectral analysis." *Analytical chemistry*, vol. 86, no. 2, pp. 1160–9, Jan. 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24354303

[99] J. L. Izquierdo-García, I. Rodríguez, A. Kyriazis, P. Villa, P. Barreiro, M. Desco, and J. Ruiz-Cabello, "A novel R-package graphic user interface for the analysis of metabonomic profiles." *BMC bioinformatics*, vol. 10, no. 1, p. 363, Jan. 2009. [Online]. Available: http://www.biomedcentral.com/1471-2105/10/363

[100] R. M. Jarvis, D. Broadhurst, H. Johnson, N. M. O'Boyle, and R. Goodacre, "PYCHEM: a multivariate analysis package for python." *Bioinformatics (Oxford, England)*, vol. 22, no. 20, pp. 2565–6, Oct. 2006. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/22/20/2565.full

[101] J. Xia, R. Mandal, I. V. Sinelnikov, D. Broadhurst, and D. S. Wishart, "MetaboAnalyst 2.0–a comprehensive server for metabolomic data analysis." *Nucleic acids research*, vol. 40, no. Web Server issue, pp. W127–33, Jul. 2012. [Online]. Available: http://nar.oxfordjournals.org/content/40/W1/W127

[102] B. Worley and R. Powers, "MVAPACK: a complete data handling package for NMR metabolomics." *ACS chemical biology*, vol. 9, no. 5, pp. 1138–44, May 2014. [Online]. Available: http://dx.doi.org/10.1021/cb4008937

[103] S. C. Booth, A. Weljie, and R. J. Turner, "Computational Tools for the Secondary Analysis of Metabolomics Experiments," Feb. 2013. [Online]. Available: http://journals.sfu.ca/rncsb/index.php/csbj/article/view/csbj.201301003/203

[104] J. C. Lindon, J. K. Nicholson, E. Holmes, H. C. Keun, A. Craig, J. T. M. Pearce, S. J. Bruce, N. Hardy, S.-A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R. D. Beger, E. R. Verheij, A. Amberg, D. Baunsgaard, G. H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J. N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I. D. Wilson, M. D. Reily, D. G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, and M. Spraul, "Summary recommendations for standardization and reporting of metabolic analyses." *Nature biotechnology*, vol. 23, no. 7, pp. 833–8, Jul. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16003371

[105] M. Chagoyen and F. Pazos, "Tools for the functional interpretation of metabolomic experiments." *Briefings in bioinformatics*, vol. 14, no. 6, pp. 737–44, Nov. 2013. [Online]. Available: http://bib.oxfordjournals.org/content/14/6/737.full

[106] E. P. Go, "Database resources in metabolomics: an overview." *Journal of neuroimmune pharmacology : the official journal of the Society*

on NeuroImmune Pharmacology, vol. 5, no. 1, pp. 18–30, Mar. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19418229

[107] J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. Velzen, J. P. M. Duijnhoven, and F. A. Dorsten, "Assessment of PLSDA cross validation," *Metabolomics*, vol. 4, no. 1, pp. 81–89, Jan. 2008. [Online]. Available: http://link.springer.com/10.1007/s11306-007-0099-6

[108] M. Andersson, "A comparison of nine PLS1 algorithms," *Journal of Chemometrics*, vol. 23, no. 10, pp. 518–529, Oct. 2009. [Online]. Available: http://doi.wiley.com/10.1002/cem.1248

[109] S. Sousa, A. Magalhães, and M. M. C. Ferreira, "Optimized bucketing for NMR spectra: Three case studies," *Chemometrics and Intelligent Laboratory Systems*, vol. 122, pp. 93–102, Mar. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743913000178

[110] P. E. Anderson, N. V. Reo, N. J. DelRaso, T. E. Doom, and M. L. Raymer, "Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics," *Metabolomics*, vol. 4, no. 3, pp. 261–272, Aug. 2008. [Online]. Available: http://corescholar.libraries.wright.edu/knoesis/83

[111] R. A. Davis, A. J. Charlton, J. Godward, S. A. Jones, M. Harrison, and J. C. Wilson, "Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 1, pp. 144–154, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743906001845

[112] P. E. Anderson, D. A. Mahle, T. E. Doom, N. V. Reo, N. J. DelRaso, and M. L. Raymer, "Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data," *Metabolomics*, vol. 7, no. 2, pp. 179–190, Nov. 2010. [Online]. Available: http://link.springer.com/10.1007/s11306-010-0242-7

[113] O. Cloarec, M. E. Dumas, J. Trygg, A. Craig, R. H. Barton, J. C. Lindon, J. K. Nicholson, and E. Holmes, "Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies." *Analytical*

chemistry, vol. 77, no. 2, pp. 517–26, Jan. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15649048

[114] G. F. Giskeø degård, T. G. Bloemberg, G. Postma, B. Sitter, M.-B. Tessem, I. S. Gribbestad, T. F. Bathen, and L. M. C. Buydens, "Alignment of high resolution magic angle spinning magnetic resonance spectra using warping methods." *Analytica chimica acta*, vol. 683, no. 1, pp. 1–11, Dec. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003267010011797

[115] F. Savorani, G. Tomasi, and S. B. Engelsen, "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra." *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 202, no. 2, pp. 190–202, Feb. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1090780709003334

[116] N.-P. V. Nielsen, J. M. Carstensen, and J. r. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping," *Journal of Chromatography A*, vol. 805, no. 1-2, pp. 17–35, May 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021967398000211

[117] G.-C. Lee and D. L. Woodruff, "Beam search for peak alignment of NMR signals," *Analytica Chimica Acta*, vol. 513, no. 2, pp. 413–416, Jun. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003267004002855

[118] D. Clifford, G. Stone, I. Montoliu, S. Rezzi, F.-P. Martin, P. Guy, S. Bruce, and S. Kochhar, "Alignment using variable penalty dynamic time warping." *Analytical chemistry*, vol. 81, no. 3, pp. 1000–7, Feb. 2009. [Online]. Available: http://dx.doi.org/10.1021/ac802041e

[119] P. H. C. Eilers, "Parametric time warping." *Analytical chemistry*, vol. 76, no. 2, pp. 404–11, Jan. 2004. [Online]. Available: http://dx.doi.org/10.1021/ac034800e

[120] G. Tomasi, F. van den Berg, and C. Andersson, "Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data," *Journal of Chemometrics*, vol. 18, no. 5, pp. 231–241, May 2004. [Online]. Available: http://doi.wiley.com/10.1002/cem.859

[121] S. Ringeissen, S. C. Connor, H. R. Brown, B. C. Sweatman, M. P. Hodson, S. P. Kenny, R. I. Haworth, P. McGill, M. A. Price, M. C. Aylott, D. J. Nunez, J. N. Haselden, and C. J. Waterfield, "Potential urinary and plasma biomarkers of peroxisome proliferation in the rat: identification of N-methylnicotinamide and N-methyl-4-pyridone-3-carboxamide by 1H nuclear magnetic resonance and high performance liquid chromatography." *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, vol. 8, no. 3-4, pp. 240–71. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12944176

[122] M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes, "NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition." *NMR in biomedicine*, vol. 18, no. 3, pp. 143–62, May 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15627238

[123] H. C. Keun, T. M. D. Ebbels, M. E. Bollard, O. Beckonert, H. Antti, E. Holmes, J. C. Lindon, and J. K. Nicholson, "Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles." *Chemical research in toxicology*, vol. 17, no. 5, pp. 579–87, May 2004. [Online]. Available: http://dx.doi.org/10.1021/tx034212w

[124] F. Dieterle, A. Ross, G. Schlotterbeck, and H. Senn, "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics." *Anal Chem*, vol. 78, no. 13, pp. 4281–4290, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1021/ac051632c

[125] D. G. Robertson, M. D. Reily, and J. D. Baker, "Metabonomics in preclinical drug development." *Expert Opin Drug Metab Toxicol*, vol. 1, no. 3, pp. 363–376, Oct. 2005. [Online]. Available: http://dx.doi.org/10.1517/17425255.1.3.363

[126] K. Pearson, "{On lines and planes of closest fit to systems of points in space}," *Philosophical Magazine*, vol. 2, no. 6, pp. 559 – 572, 1901.

[127] H. Hotelling, "Analysis of a complex of statistical variables into principal components," Baltimore :, 1933.

[128] I. Jolliffe, *Principal Component Analysis*. Springer, 2002. [Online].

Available: http://books.google.co.uk/books/about/Principal_Component_Analysis.html?id=_olByCrhjwIC&pgis=1

[129] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, Oct. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743901001551

[130] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics*, vol. 16, no. 3, pp. 119–128, 2002. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/cem.695/abstracthttp://onlinelibrary.wiley.com/store/10.1002/cem.695/asset/695_ftp.pdf?v=1%&t=h6fa8amq&s=53fcbe01ce32d39ac5a7bda8715c4c4ac3b6d8db

[131] H. S. Tapp and E. K. Kemsley, "Notes on the practical utility of OPLS," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 11, pp. 1322–1327, Dec. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.trac.2009.08.006

[132] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982. [Online]. Available: http://link.springer.com/10.1007/BF00337288

[133] A. I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush, "TM4: a free, open-source system for microarray data management and analysis." *Biotechniques*, vol. 34, no. 2, pp. 374–378, Feb. 2003.

[134] R. Bro, K. Kjeldahl, A. K. Smilde, and H. A. L. Kiers, "Cross-validation of component models: a critical look at current methods." *Analytical and bioanalytical chemistry*, vol. 390, no. 5, pp. 1241–51, Mar. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18214448

[135] C.-J. Xu, H. C. Hoefsloot, and A. K. Smilde, "To aggregate or not to aggregate high-dimensional classifiers." *BMC Bioinformatics*, vol. 12, p. 153, 2011. [Online]. Available: http://dx.doi.org/10.1186/1471-2105-12-153

[136] B. J. A. Mertens, M. E. D. Noo, R. A. E. M. Tollenaar, and A. M. Deelder, "Mass spectrometry proteomic diagnosis: enacting the double cross-validatory paradigm." *J Comput Biol*, vol. 13, no. 9, pp. 1591–1605, Nov. 2006. [Online]. Available: http://dx.doi.org/10.1089/cmb.2006.13.1591

[137] E. Szymaska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, "Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies." *Metabolomics : Official journal of the Metabolomic Society*, vol. 8, no. Suppl 1, pp. 3–16, Jun. 2012. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3337399&tool=pmcentrez&rendertype=abstract

[138] W. Haller, O. Ledder, P. J. Lewindon, R. Couper, K. J. Gaskin, and M. Oliver, "Cystic fibrosis: An update for clinicians. Part 1: Nutrition and gastrointestinal complications," *Journal of Gastroenterology and Hepatology*, vol. 29, no. 7, pp. 1344–1355, Jul. 2014. [Online]. Available: http://doi.wiley.com/10.1111/jgh.12546

[139] Cystic Fibrosis Trust, "UK Cystic Fibrosis Registry Annual data report 2013," no. July, 2014.

[140] P. R. Sosnay, K. R. Siklosi, F. Van Goor, K. Kaniecki, H. Yu, N. Sharma, A. S. Ramalho, M. D. Amaral, R. Dorfman, J. Zielenski, D. L. Masica, R. Karchin, L. Millen, P. J. Thomas, G. P. Patrinos, M. Corey, M. H. Lewis, J. M. Rommens, C. Castellani, C. M. Penland, and G. R. Cutting, "Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene." *Nature genetics*, vol. 45, no. 10, pp. 1160–7, Oct. 2013. [Online]. Available: http://dx.doi.org/10.1038/ng.2745

[141] J. B. Lyczak, C. L. Cannon, and G. B. Pier, "Lung infections associated with cystic fibrosis." *Clinical microbiology reviews*, vol. 15, no. 2, pp. 194–222, Apr. 2002. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=118069&tool=pmcentrez&rendertype=abstract

[142] S. C. FitzSimmons, "The changing epidemiology of cystic fibrosis," *The Journal of Pediatrics*, vol. 122, no. 1, pp. 1–9, Jan. 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002234760583478X

[143] P. Greally, P. Whitaker, and D. Peckham, "Challenges with current inhaled treatments for chronic Pseudomonas aeruginosa infection in

patients with cystic fibrosis." *Current medical research and opinion*, vol. 28, no. 6, pp. 1059–67, Jun. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22401602

[144] G. Döring, S. P. Conway, H. G. Heijerman, M. E. Hodson, N. Hø iby, A. Smyth, and D. J. Touw, "Antibiotic therapy against Pseudomonas aeruginosa in cystic fibrosis: a European consensus." *The European respiratory journal*, vol. 16, no. 4, pp. 749–67, Oct. 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11106223

[145] R. L. Gibson, J. L. Burns, and B. W. Ramsey, "Pathophysiology and management of pulmonary infections in cystic fibrosis." *American journal of respiratory and critical care medicine*, vol. 168, no. 8, pp. 918–51, Oct. 2003. [Online]. Available: http://www.atsjournals.org/doi/full/10.1164/rccm.200304-505SO#.VBAZUPldUrV

[146] K. M. Noone PG, "Standard therapy of CF lung disease," in *Cystic fibrosis in adults*, 1999, pp. 145–175.

[147] G. B. Rogers, M. P. Carroll, D. J. Serisier, P. M. Hockey, G. Jones, V. Kehagia, G. J. Connett, and K. D. Bruce, "Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis." *Journal of clinical microbiology*, vol. 44, no. 7, pp. 2601–4, Jul. 2006. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1489498&tool=pmcentrez&rendertype=abstract

[148] G. B. Rogers, F. A. Stressmann, A. W. Walker, M. P. Carroll, and K. D. Bruce, "Lung infections in cystic fibrosis: deriving clinical insight from microbial complexity." *Expert review of molecular diagnostics*, vol. 10, no. 2, pp. 187–96, Mar. 2010. [Online]. Available: http://informahealthcare.com/doi/abs/10.1586/erm.09.81

[149] O. Ciofu, L. F. Mandsberg, H. Wang, and N. Hø iby, "Phenotypes selected during chronic lung infection in cystic fibrosis patients: implications for the treatment of Pseudomonas aeruginosa biofilm infections." *FEMS immunology and medical microbiology*, vol. 65, no. 2, pp. 215–25, Jul. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22540844

[150] V. Klepac-Ceraj, K. P. Lemon, T. R. Martin, M. Allgaier, S. W. Kembel, A. A. Knapp, S. Lory, E. L. Brodie, S. V. Lynch, B. J. M. Bohannan, J. L. Green, B. A. Maurer, and R. Kolter, "Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa." *Environmental microbiology*, vol. 12, no. 5, pp. 1293–303, May 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20192960

[151] C. J. van der Gast, A. W. Walker, F. A. Stressmann, G. B. Rogers, P. Scott, T. W. Daniels, M. P. Carroll, J. Parkhill, and K. D. Bruce, "Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities." *The ISME journal*, vol. 5, no. 5, pp. 780–91, May 2011. [Online]. Available: http://dx.doi.org/10.1038/ismej.2010.175

[152] A. A. Fodor, E. R. Klem, D. F. Gilpin, J. S. Elborn, R. C. Boucher, M. M. Tunney, and M. C. Wolfgang, "The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations." *PloS one*, vol. 7, no. 9, p. e45001, Jan. 2012. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0045001

[153] E. Kerem, J. Reisman, M. Corey, G. J. Canny, and H. Levison, "Prediction of mortality in patients with cystic fibrosis." *The New England journal of medicine*, vol. 326, no. 18, pp. 1187–91, Apr. 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1285737

[154] M. B. Butterworth, L. Zhang, E. M. Heidrich, M. M. Myerburg, and P. H. Thibodeau, "Activation of the epithelial sodium channel (ENaC) by the alkaline protease from Pseudomonas aeruginosa." *The Journal of biological chemistry*, vol. 287, no. 39, pp. 32 556–65, Sep. 2012. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3463336&tool=pmcentrez&rendertype=abstract

[155] L. McGuigan and M. Callaghan, "The evolving dynamics of the microbial community in the cystic fibrosis lung." *Environmental microbiology*, May 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24801013

[156] E. L. Gjersing, J. L. Herberg, J. Horn, C. M. Schaldach, and R. S. Maxwell, "NMR metabolomics of planktonic and biofilm modes of growth in Pseudomonas aeruginosa." *Analytical chemistry*,

vol. 79, no. 21, pp. 8037–45, Nov. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17915964

[157] M. M. Feizabadi, A. Majnooni, B. Nomanpour, B. Fatolahzadeh, N. Raji, S. Delfani, M. Habibi, S. Asadi, and M. Parvin, "Direct detection of Pseudomonas aeruginosa from patients with healthcare associated pneumonia by real time PCR." *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, vol. 10, no. 8, pp. 1247–51, Dec. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20728580

[158] N. Renders, Y. Römling, H. Verbrugh, and A. van Belkum, "Comparative typing of Pseudomonas aeruginosa by random amplification of polymorphic DNA or pulsed-field gel electrophoresis of DNA macrorestriction fragments." *Journal of clinical microbiology*, vol. 34, no. 12, pp. 3190–5, Dec. 1996. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=229481&tool=pmcentrez&rendertype=abstract

[159] D. D. Sriramulu, H. Lünsdorf, J. S. Lam, and U. Römling, "Microcolony formation: a novel biofilm model of Pseudomonas aeruginosa for the cystic fibrosis lung." *Journal of medical microbiology*, vol. 54, no. Pt 7, pp. 667–76, Jul. 2005. [Online]. Available: http://jmm.sgmjournals.org/content/54/7/667.abstract

[160] C. Palmer, E. M. Bik, D. B. DiGiulio, D. A. Relman, and P. O. Brown, "Development of the human infant intestinal microbiota." *PLoS biology*, vol. 5, no. 7, p. e177, Jul. 2007. [Online]. Available: http://dx.plos.org/10.1371/journal.pbio.0050177

[161] K. L. Palmer, L. M. Mashburn, P. K. Singh, and M. Whiteley, "Cystic fibrosis sputum supports growth and cues key aspects of Pseudomonas aeruginosa physiology." *Journal of bacteriology*, vol. 187, no. 15, pp. 5267–77, Aug. 2005. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1196007&tool=pmcentrez&rendertype=abstract

[162] C. Fung, S. Naughton, L. Turnbull, P. Tingpej, B. Rose, J. Arthur, H. Hu, C. Harmer, C. Harbour, D. J. Hassett, C. B. Whitchurch, and J. Manos, "Gene expression of Pseudomonas aeruginosa in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum." *Journal*

of medical microbiology, vol. 59, no. Pt 9, pp. 1089–100, Sep. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20522626

[163] Y. Mitsui, K. Matsumura, C. Kondo, and R. Takashima, "The role of mucin on experimental Pseudomonas keratitis in rabbits." *Investigative ophthalmology*, vol. 15, no. 3, pp. 208–10, Mar. 1976. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/815195

[164] T. Hwang and A. Shaka, "Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients," *Journal of Magnetic Resonance, Series A*, vol. 112, no. 2, pp. 275–279, Feb. 1995. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1064185885710479

[165] M. S. Son, W. J. Matthews, Y. Kang, D. T. Nguyen, and T. T. Hoang, "In vivo evidence of Pseudomonas aeruginosa nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients." *Infection and immunity*, vol. 75, no. 11, pp. 5313–24, Nov. 2007. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2168270&tool=pmcentrez&rendertype=abstract

[166] E. Frimmersdorf, S. Horatzek, A. Pelnikevich, L. Wiehlmann, and D. Schomburg, "How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach." *Environmental microbiology*, vol. 12, no. 6, pp. 1734–47, Jun. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20553553

[167] A. Beneduci, G. Chidichimo, G. Dardo, and G. Pontoni, "Highly routinely reproducible alignment of 1H NMR spectral peaks of metabolites in huge sets of urines." *Analytica chimica acta*, vol. 685, no. 2, pp. 186–95, Jan. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21168568

[168] H. Resat, V. Bailey, L. A. McCue, and A. Konopka, "Modeling microbial dynamics in heterogeneous environments: growth on soil carbon sources." *Microbial ecology*, vol. 63, no. 4, pp. 883–97, May 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22193925

[169] S. J. Dunaj, J. J. Vallino, M. E. Hines, M. Gay, C. Kobyljanec, and J. N. Rooney-Varga, "Relationships between soil organic matter, nutrients, bacterial community structure, and the performance of microbial fuel

cells." *Environmental science & technology*, vol. 46, no. 3, pp. 1914–22, Feb. 2012. [Online]. Available: http://dx.doi.org/10.1021/es2032532

[170] D. Nguyen, A. Joshi-Datar, F. Lepine, E. Bauerle, O. Olakanmi, K. Beer, G. McKay, R. Siehnel, J. Schafhauser, Y. Wang, B. E. Britigan, and P. K. Singh, "Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria." *Science (New York, N.Y.)*, vol. 334, no. 6058, pp. 982–6, Nov. 2011. [Online]. Available: http://www.sciencemag.org/content/334/6058/982

[171] S. P. Bernier, D.-G. Ha, W. Khan, J. H. Merritt, and G. A. O'Toole, "Modulation of Pseudomonas aeruginosa surface-associated group behaviors by individual amino acids through c-di-GMP signaling." *Research in microbiology*, vol. 162, no. 7, pp. 680–8, Sep. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3716369&tool=pmcentrez&rendertype=abstract

[172] J. D. Shrout, D. L. Chopp, C. L. Just, M. Hentzer, M. Givskov, and M. R. Parsek, "The impact of quorum sensing and swarming motility on Pseudomonas aeruginosa biofilm formation is nutritionally conditional." *Molecular microbiology*, vol. 62, no. 5, pp. 1264–77, Dec. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17059568

[173] K. Romanowski, A. Zaborin, H. Fernandez, V. Poroyko, V. Valuckaite, S. Gerdes, D. C. Liu, O. Y. Zaborina, and J. C. Alverdy, "Prevention of siderophore- mediated gut-derived sepsis due to P. aeruginosa can be achieved without iron provision by maintaining local phosphate abundance: role of pH." *BMC microbiology*, vol. 11, no. 1, p. 212, Jan. 2011. [Online]. Available: http://www.biomedcentral.com/1471-2180/11/212

[174] S. H. Duncan, P. Louis, J. M. Thomson, and H. J. Flint, "The role of pH in determining the species composition of the human colonic microbiota." *Environmental microbiology*, vol. 11, no. 8, pp. 2112–22, Aug. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19397676

[175] A. W. Walker, S. H. Duncan, E. C. McWilliam Leitch, M. W. Child, and H. J. Flint, "pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon." *Applied and environmental*

microbiology, vol. 71, no. 7, pp. 3692–700, Jul. 2005. [Online]. Available: http://aem.asm.org/content/71/7/3692

[176] C. Clary-Meinesz, J. Mouroux, J. Cosson, P. Huitorel, and B. Blaive, "Influence of external pH on ciliary beat frequency in human bronchi and bronchioles." *The European respiratory journal*, vol. 11, no. 2, pp. 330–3, Feb. 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9551733

[177] S. K. Inglis, M. R. Corboz, and S. T. Ballard, "Effect of anion secretion inhibitors on mucin content of airway submucosal gland ducts." *The American journal of physiology*, vol. 274, no. 5 Pt 1, pp. L762–6, May 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9612291

[178] T. G. Kloosterman and O. P. Kuipers, "Regulation of arginine acquisition and virulence gene expression in the human pathogen Streptococcus pneumoniae by transcription regulators ArgR1 and AhrC." *The Journal of biological chemistry*, vol. 286, no. 52, pp. 44 594–605, Dec. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3248006&tool=pmcentrez&rendertype=abstract

[179] M. Rogosa and F. S. Bishop, "THE GENUS VEILLONELLA . II. Nutritional Studies." *Journal of bacteriology*, vol. 87, pp. 574–80, Mar. 1964. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=277056&tool=pmcentrez&rendertype=abstract

[180] The Genome Institute, "The Human Gut Microbiome Initiative (HGMI)," Tech. Rep., 2013. [Online]. Available: http://genome.wustl.edu/projects/detail/human-gut-microbiome/

[181] P. B. Eckburg, E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman, "Diversity of the human intestinal microbial flora." *Science (New York, N.Y.)*, vol. 308, no. 5728, pp. 1635–8, Jun. 2005. [Online]. Available: http://www.sciencemag.org/content/308/5728/1635

[182] S. Matamoros, C. Gras-Leguen, F. Le Vacon, G. Potel, and M.-F. de La Cochetiere, "Development of intestinal microbiota in infants and its impact on health." *Trends in microbiology*, vol. 21, no. 4, pp. 167–73, Apr. 2013. [Online]. Available: http://www.cell.com/article/S0966842X12002132/fulltext

[183] J. L. Round and S. K. Mazmanian, "The gut microbiota shapes intestinal immune responses during health and disease." *Nature reviews. Immunology*, vol. 9, no. 5, pp. 313–23, May 2009. [Online]. Available: http://dx.doi.org/10.1038/nri2515

[184] R. Stepankova, Z. Tonar, J. Bartova, L. Nedorost, P. Rossman, R. Poledne, M. Schwarzer, and H. Tlaskalova-Hogenova, "Absence of microbiota (germ-free conditions) accelerates the atherosclerosis in ApoE-deficient mice fed standard low cholesterol diet." *Journal of atherosclerosis and thrombosis*, vol. 17, no. 8, pp. 796–804, Aug. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20379054

[185] F. Bäckhed, "Changes in intestinal microflora in obesity: cause or consequence?" *Journal of pediatric gastroenterology and nutrition*, vol. 48 Suppl 2, pp. S56–7, Apr. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19300127

[186] P. D. Cani, R. Bibiloni, C. Knauf, A. Waget, A. M. Neyrinck, N. M. Delzenne, and R. Burcelin, "Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice." *Diabetes*, vol. 57, no. 6, pp. 1470–81, Jun. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18305141

[187] M. Vijay-Kumar, J. D. Aitken, F. A. Carvalho, T. C. Cullender, S. Mwangi, S. Srinivasan, S. V. Sitaraman, R. Knight, R. E. Ley, and A. T. Gewirtz, "Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5." *Science (New York, N.Y.)*, vol. 328, no. 5975, pp. 228–31, Apr. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20203013

[188] D. N. Frank, A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 34, pp. 13 780–5, Aug. 2007. [Online]. Available: http://www.pnas.org/content/104/34/13780.abstract

[189] B. Björkholm, C. M. Bok, A. Lundin, J. Rafter, M. L. Hibberd, and S. Pettersson, "Intestinal microbiota regulate xenobiotic metabolism in the liver." *PloS one*, vol. 4, no. 9, p. e6958, Jan. 2009. [Online].

Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2734986&tool=pmcentrez&rendertype=abstract

[190] R. Diaz Heijtz, S. Wang, F. Anuar, Y. Qian, B. Björkholm, A. Samuelsson, M. L. Hibberd, H. Forssberg, and S. Pettersson, "Normal gut microbiota modulates brain development and behavior." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 7, pp. 3047–52, Feb. 2011. [Online]. Available: http://www.pnas.org/content/early/2011/01/26/1010529108.short

[191] S. H. Lee, L.-L. Hu, J. Gonzalez-Navajas, G. S. Seo, C. Shen, J. Brick, S. Herdman, N. Varki, M. Corr, J. Lee, and E. Raz, "ERK activation drives intestinal tumorigenesis in Apc(min/+) mice." *Nature medicine*, vol. 16, no. 6, pp. 665–70, Jun. 2010. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2882530&tool=pmcentrez&rendertype=abstract

[192] I. Cho and M. J. Blaser, "The human microbiome: at the interface of health and disease." *Nature reviews. Genetics*, vol. 13, no. 4, pp. 260–70, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1038/nrg3182

[193] I. Sekirov, S. L. Russell, L. C. M. Antunes, and B. B. Finlay, "Gut microbiota in health and disease." *Physiological reviews*, vol. 90, no. 3, pp. 859–904, Jul. 2010. [Online]. Available: http://physrev.physiology.org/content/90/3/859.short

[194] V. Tremaroli and F. Bäckhed, "Functional interactions between the gut microbiota and host metabolism." *Nature*, vol. 489, no. 7415, pp. 242–9, Sep. 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11552

[195] J. K. Nicholson, E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia, and S. Pettersson, "Host-gut microbiota metabolic interactions." *Science (New York, N.Y.)*, vol. 336, no. 6086, pp. 1262–7, Jun. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22674330

[196] A. Yoshiki and K. Moriwaki, "Mouse Phenome Research: Implications of Genetic Background," *ILAR Journal*, vol. 47, no. 2, pp. 94–102, Jan. 2006. [Online]. Available: http://ilarjournal.oxfordjournals.org/content/47/2/94.long

[197] A. Marcobal, P. C. Kashyap, T. A. Nelson, P. A. Aronov, M. S. Donia, A. Spormann, M. A. Fischbach, and J. L. Sonnenburg, "A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice." *The ISME journal*, vol. 7, no. 10, pp. 1933–43, Oct. 2013. [Online]. Available: http://dx.doi.org/10.1038/ismej.2013.89

[198] M. Matsumoto, R. Kibe, T. Ooga, Y. Aiba, S. Kurihara, E. Sawaki, Y. Koga, and Y. Benno, "Impact of intestinal microbiota on intestinal luminal metabolome." *Scientific reports*, vol. 2, p. 233, Jan. 2012. [Online]. Available: http://www.nature.com/srep/2012/120125/srep00233/full/srep00233.html

[199] B. Haegeman, J. Hamelin, J. Moriarty, P. Neal, J. Dushoff, and J. S. Weitz, "Robust estimation of microbial diversity in theory and in practice." *The ISME journal*, vol. 7, no. 6, pp. 1092–101, Jun. 2013. [Online]. Available: http://dx.doi.org/10.1038/ismej.2013.10

[200] M. Linnenbrink, J. Wang, E. A. Hardouin, S. Künzel, D. Metzler, and J. F. Baines, "The role of biogeography in shaping diversity of the intestinal microbiota in house mice." *Molecular ecology*, vol. 22, no. 7, pp. 1904–16, Apr. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23398547

[201] M. Roberfroid, G. R. Gibson, L. Hoyles, A. L. McCartney, R. Rastall, I. Rowland, D. Wolvers, B. Watzl, H. Szajewska, B. Stahl, F. Guarner, F. Respondek, K. Whelan, V. Coxam, M.-J. Davicco, L. Léotoing, Y. Wittrant, N. M. Delzenne, P. D. Cani, A. M. Neyrinck, and A. Meheust, "Prebiotic effects: metabolic and health benefits." *The British journal of nutrition*, vol. 104 Suppl, pp. S1–63, Aug. 2010. [Online]. Available: http://journals.cambridge.org/action/displayFulltext?type=6&fid=7910382&jid=BJN&volumeId=104&issueId=S2&aid=7910381&bodyId=&membershipNumber=%&societyETOCSession=&fulltextType=RA&fileId=S0007114510003363

[202] N. M. Delzenne and P. D. Cani, "Interaction between obesity and the gut microbiota: relevance in nutrition." *Annual review of nutrition*, vol. 31, pp. 15–31, Aug. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21568707

[203] M. Blaut, "Relationship of prebiotics and food to intestinal microflora." *European journal of nutrition*, vol. 41 Suppl 1, pp. I11–6, Nov. 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12420111

[204] K. M. Maslowski, A. T. Vieira, A. Ng, J. Kranich, F. Sierro, D. Yu, H. C. Schilter, M. S. Rolph, F. Mackay, D. Artis, R. J. Xavier, M. M. Teixeira, and C. R. Mackay, "Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43." *Nature*, vol. 461, no. 7268, pp. 1282–6, Oct. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3256734&tool=pmcentrez&rendertype=abstract

[205] A. Marcobal, P. C. Kashyap, T. A. Nelson, P. A. Aronov, M. S. Donia, A. Spormann, M. A. Fischbach, and J. L. Sonnenburg, "A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice." *The ISME journal*, vol. 7, no. 10, pp. 1933–43, Oct. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23739052

[206] L. K. Ursell, H. J. Haiser, W. Van Treuren, N. Garg, L. Reddivari, J. Vanamala, P. C. Dorrestein, P. J. Turnbaugh, and R. Knight, "The intestinal metabolome: an intersection between microbiota and host." *Gastroenterology*, vol. 146, no. 6, pp. 1470–6, May 2014. [Online]. Available: http://www.gastrojournal.org/article/S0016508514002996/fulltext

[207] W. R. Wikoff, A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters, and G. Siuzdak, "Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 10, pp. 3698–703, Mar. 2009. [Online]. Available: http://www.pnas.org/content/early/2009/02/19/0812874106

[208] F.-P. J. Martin, N. Sprenger, I. K. S. Yap, Y. Wang, R. Bibiloni, F. Rochat, S. Rezzi, C. Cherbut, S. Kochhar, J. C. Lindon, E. Holmes, and J. K. Nicholson, "Panorganismal gut microbiome-host metabolic crosstalk." *Journal of proteome research*, vol. 8, no. 4, pp. 2090–105, Apr. 2009. [Online]. Available: http://dx.doi.org/10.1021/pr801068x

[209] L. S. Vermeer, G. O. Fruhwirth, P. Pandya, T. Ng, and A. J. Mason, "NMR metabolomics of MTLn3E breast cancer cells identifies

a role for CXCR4 in lipid and choline regulation." *Journal of proteome research*, vol. 11, no. 5, pp. 2996–3003, May 2012. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3378657&tool=pmcentrez&rendertype=abstract

[210] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. r. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. r. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, S. D. Ehrlich, and J. Wang, "A human gut microbial gene catalogue established by metagenomic sequencing." *Nature*, vol. 464, no. 7285, pp. 59–65, Mar. 2010. [Online]. Available: http://dx.doi.org/10.1038/nature08821

[211] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. r. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. r. Brunak, J. Doré, M. Antolín, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C. Cara, C. Chervaux, A. Cultrone, C. Delorme, G. Denariaz, R. Dervyn, K. U. Foerstner, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg, A. Jamet, C. Juste, G. Kaci, J. Knol, O. Lakhdari, S. Layec, K. Le Roux, E. Maguin, A. Mérieux, R. Melo Minardi, C. M'rini, J. Muller, R. Oozeer, J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K. Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J. Weissenbach, S. D. Ehrlich, and P. Bork, "Enterotypes of the human gut microbiome." *Nature*, vol. 473, no. 7346, pp. 174–80, May 2011. [Online]. Available: http://dx.doi.org/10.1038/nature09944

[212] The Human Microbiome Project (HMP), "Structure, function and diversity of the healthy human microbiome." *Nature*, vol. 486, no. 7402, pp. 207–14, Jun. 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11234

[213] P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, and J. I. Gordon, "A core gut microbiome in obese and lean twins." *Nature*, vol. 457, no. 7228, pp. 480–4, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1038/nature07540

[214] I. H. McHardy, M. Goudarzi, M. Tong, P. M. Ruegger, E. Schwager, J. R. Weger, T. G. Graeber, J. L. Sonnenburg, S. Horvath, C. Huttenhower, D. P. McGovern, A. J. Fornace, J. Borneman, and J. Braun, "Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships." *Microbiome*, vol. 1, no. 1, p. 17, Jan. 2013. [Online]. Available: http://www.microbiomejournal.com/content/1/1/17

[215] P. J. Turnbaugh, V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight, and J. I. Gordon, "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice." *Science translational medicine*, vol. 1, no. 6, p. 6ra14, Nov. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2894525&tool=pmcentrez&rendertype=abstract

[216] G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Li, F. D. Bushman, and J. D. Lewis, "Linking long-term dietary patterns with gut microbial enterotypes." *Science (New York, N.Y.)*, vol. 334, no. 6052, pp. 105–8, Oct. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3368382&tool=pmcentrez&rendertype=abstract

[217] M. J. Claesson, I. B. Jeffery, S. Conde, S. E. Power, E. M. O'Connor, S. Cusack, H. M. B. Harris, M. Coakley, B. Lakshminarayanan, O. O'Sullivan, G. F. Fitzgerald, J. Deane, M. O'Connor, N. Harnedy, K. O'Connor, D. O'Mahony, D. van Sinderen, M. Wallace, L. Brennan, C. Stanton, J. R. Marchesi, A. P. Fitzgerald, F. Shanahan, C. Hill, R. P. Ross, and P. W. O'Toole, "Gut microbiota composition correlates with diet and health in the elderly." *Nature*, vol. 488, no. 7410, pp. 178–84, Aug. 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11319

[218] A. M. Allahverdiyev, M. Bagirova, E. S. Abamor, S. C. Ates, R. C. Koc, M. Miraloglu, S. Elcicek, S. Yaman, and G. Unal,

"The use of platensimycin and platencin to fight antibiotic resistance." *Infection and drug resistance*, vol. 6, pp. 99–114, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3785399&tool=pmcentrez&rendertype=abstract

[219] P.-R. Hsueh, W.-H. Chen, and K.-T. Luh, "Relationships between antimicrobial use and antimicrobial resistance in Gram-negative bacteria causing nosocomial infections from 1991-2003 at a university hospital in Taiwan." *International journal of antimicrobial agents*, vol. 26, no. 6, pp. 463–72, Dec. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16280243

[220] B. A. Cunha, "Antibiotic resistance." *The Medical clinics of North America*, vol. 84, no. 6, pp. 1407–29, Nov. 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11155850

[221] I.-L. Chen, C.-H. Lee, L.-H. Su, Y.-F. Tang, S.-J. Chang, and J.-W. Liu, "Antibiotic consumption and healthcare-associated infections caused by multidrug-resistant gram-negative bacilli at a large medical center in Taiwan from 2002 to 2009: implicating the importance of antibiotic stewardship." *PloS one*, vol. 8, no. 5, p. e65621, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3667806&tool=pmcentrez&rendertype=abstract

[222] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, and J. Bartlett, "Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America." *Clin Infect Dis*, vol. 48, no. 1, pp. 1–12, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1086/595011

[223] Centers for Disease Control and Prevention, "Antibiotic Resistance threats in the United States, 2013," Tech. Rep., 2013.

[224] H. W. Boucher and G. R. Corey, "Epidemiology of methicillin-resistant Staphylococcus aureus." *Clin Infect Dis*, vol. 46 Suppl 5, pp. S344—-S349, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1086/533590

[225] M. E. Falagas and I. A. Bliziotis, "Pandrug-resistant Gram-negative bacteria: the dawn of the post-antibiotic era?" *Int J Antimicrob Agents*, vol. 29, no. 6, pp. 630–636, Jun. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.ijantimicag.2006.12.012

[226] S. R. Palumbi, "Humans as the world's greatest evolutionary force." *Science*, vol. 293, no. 5536, pp. 1786–1790, Sep. 2001. [Online]. Available: http://dx.doi.org/10.1126/science.293.5536.1786

[227] M. S. Butler, M. A. Blaskovich, and M. A. Cooper, "Antibiotics in the clinical pipeline in 2013." *The Journal of antibiotics*, vol. 66, no. 10, pp. 571–91, Oct. 2013. [Online]. Available: http://dx.doi.org/10.1038/ja.2013.86

[228] R. O'Shea and H. E. Moser, "Physicochemical properties of antibacterial compounds: implications for drug discovery." *Journal of medicinal chemistry*, vol. 51, no. 10, pp. 2871–8, May 2008. [Online]. Available: http://dx.doi.org/10.1021/jm700967e

[229] L. L. Silver, "Challenges of antibacterial discovery." *Clinical microbiology reviews*, vol. 24, no. 1, pp. 71–109, Jan. 2011. [Online]. Available: http://cmr.asm.org/content/24/1/71

[230] S. Davies, "Chief Medical Officer annual report: volume 2 - Publications - GOV.UK," Tech. Rep., 2013. [Online]. Available: https://www.gov.uk/government/publications/chief-medical-officer-annual-report-volume-2

[231] J. Davies, "Inactivation of antibiotics and the dissemination of resistance genes," *Science*, vol. 264, no. 5157, pp. 375–382, Apr. 1994. [Online]. Available: http://www.sciencemag.org/content/264/5157/375

[232] M. K. Waldor and D. I. Friedman, "Phage regulatory circuits and virulence gene expression." *Current opinion in microbiology*, vol. 8, no. 4, pp. 459–65, Aug. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15979389

[233] D. W. Acheson and A. Donohue-Rolfe, "Cancer-associated hemolytic uremic syndrome: a possible role of mitomycin in relation to Shiga-like toxins." *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 7, no. 12, p. 1943, Dec. 1989. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2511278

[234] P. L. Wagner, J. Livny, M. N. Neely, D. W. K. Acheson, D. I. Friedman, and M. K. Waldor, "Bacteriophage control of Shiga toxin 1 production and release by Escherichia coli." *Molecular microbiology*, vol. 44, no. 4, pp. 957–70, May 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12010491

[235] P. T. Kimmitt, C. R. Harwood, and M. R. Barer, "Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response." *Emerging infectious diseases*, vol. 6, no. 5, pp. 458–65. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2627954&tool=pmcentrez&rendertype=abstract

[236] E. Guerin, G. Cambray, N. Sanchez-Alberola, S. Campoy, I. Erill, S. Da Re, B. Gonzalez-Zorn, J. Barbé, M.-C. Ploy, and D. Mazel, "The SOS response controls integron recombination." *Science (New York, N.Y.)*, vol. 324, no. 5930, p. 1034, May 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19460999

[237] C. Miller, L. E. Thomsen, C. Gaggero, R. Mosseri, H. Ingmer, and S. N. Cohen, "SOS response induction by beta-lactams and bacterial defense against antibiotic lethality." *Science (New York, N.Y.)*, vol. 305, no. 5690, pp. 1629–31, Sep. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15308764

[238] J. Wu, Q. Long, and J. Xie, "(p)ppGpp and drug resistance." *Journal of cellular physiology*, vol. 224, no. 2, pp. 300–4, Aug. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20432457

[239] D. G. Rodionov and E. E. Ishiguro, "Direct correlation between overproduction of guanosine 3',5'-bispyrophosphate (ppGpp) and penicillin tolerance in Escherichia coli." *Journal of bacteriology*, vol. 177, no. 15, pp. 4224–9, Aug. 1995. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=177166&tool=pmcentrez&rendertype=abstract

[240] S. W. Mamber, B. Kolek, K. W. Brookshire, D. P. Bonner, and J. Fung-Tomc, "Activity of quinolones in the Ames Salmonella TA102 mutagenicity test and other bacterial genotoxicity assays." *Antimicrobial agents and chemotherapy*, vol. 37, no. 2, pp. 213–7, Feb. 1993. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=187641&tool=pmcentrez&rendertype=abstract

[241] R. T. Cirz, J. K. Chin, D. R. Andes, V. de Crécy-Lagard, W. A. Craig, and F. E. Romesberg, "Inhibition of mutation and combating the evolution of antibiotic resistance." *PLoS biology*, vol. 3, no. 6, p. e176, Jun. 2005. [Online]. Available: http://dx.plos.org/10.1371/journal.pbio.0030176

[242] J. W. Beaber, B. Hochhut, and M. K. Waldor, "SOS response promotes horizontal dissemination of antibiotic resistance genes." *Nature*, vol. 427, no. 6969, pp. 72–4, Jan. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14688795

[243] R. A. F. Wozniak, D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, G. Garriss, C. Déry, V. Burrus, and M. K. Waldor, "Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs." *PLoS genetics*, vol. 5, no. 12, p. e1000786, Dec. 2009. [Online]. Available: http://dx.plos.org/10.1371/journal.pgen.1000786

[244] S. Balashov and M. Z. Humayun, "Escherichia coli cells bearing a ribosomal ambiguity mutation in rpsD have a mutator phenotype that correlates with increased mistranslation." *Journal of bacteriology*, vol. 185, no. 16, pp. 5015–8, Aug. 2003. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=166475&tool=pmcentrez&rendertype=abstract

[245] K. Poole, "Bacterial Multidrug Efflux Pumps Serve Other Functions," *Microbes*, vol. 3, no. 4, pp. 179–185, 2008. [Online]. Available: http://www.microbemagazine.org/index.php?option=com_content&view=article&id=969:bacterial-multidrug-efflux-pumps-serve-other-functions&catid=293%&Itemid=431

[246] Z. D. Dalebroux, S. Matamouros, D. Whittington, R. E. Bishop, and S. I. Miller, "PhoPQ regulates acidic glycerophospholipid content of the Salmonella Typhimurium outer membrane." *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1 316 901 111–, Jan. 2014. [Online]. Available: http://www.pnas.org/content/early/2014/01/17/1316901111.abstract

[247] H. S. Girgis, A. K. Hottes, and S. Tavazoie, "Genetic architecture of intrinsic antibiotic susceptibility." *PloS one*, vol. 4, no. 5, p. e5629, Jan. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2680486&tool=pmcentrez&rendertype=abstract

[248] M. A. Kohanski, M. A. DePristo, and J. J. Collins, "Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis." *Molecular cell*, vol. 37, no. 3, pp. 311–20, Feb. 2010. [Online].

Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2840266&tool=pmcentrez&rendertype=abstract

[249] O. F. Join-Lambert, M. Michéa-Hamzehpour, T. Köhler, F. Chau, F. Faurisson, S. Dautrey, C. Vissuzaine, C. Carbon, and J. Pechère, "Differential selection of multidrug efflux mutants by trovafloxacin and ciprofloxacin in an experimental model of Pseudomonas aeruginosa acute pneumonia in rats." *Antimicrobial agents and chemotherapy*, vol. 45, no. 2, pp. 571–6, Feb. 2001. [Online]. Available: http://aac.asm.org/content/45/2/571.abstract?ijkey=9651180d2bb3da109d3118ad52a28c97f69d0351&keytype2=tf_ipsecsha

[250] R. H. Eng, F. T. Padberg, S. M. Smith, E. N. Tan, and C. E. Cherubin, "Bactericidal effects of antibiotics on slowly growing and nongrowing bacteria." *Antimicrobial Agents and Chemotherapy*, vol. 35, no. 9, pp. 1824–1828, Sep. 1991. [Online]. Available: http://aac.asm.org/content/35/9/1824.abstract?ijkey=55b10f486689fbd59b8ccac4696559adf4e168b8&keytype2=tf_ipsecsha

[251] E. E. Ishiguro and W. D. Ramey, "Inhibition of in vitro peptidoglycan biosynthesis in Escherichia coli by guanosine 5'-diphosphate 3'-diphosphate." *Canadian journal of microbiology*, vol. 26, no. 12, pp. 1514–8, Dec. 1980. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/7016279

[252] P. Landini, "Cross-talk mechanisms in biofilm formation and responses to environmental and physiological stress in Escherichia coli," *Research in Microbiology*, vol. 160, no. 4, pp. 259–266, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923250809000230

[253] A. Sandiumenge, E. Diaz, A. Rodriguez, L. Vidaur, L. Canadell, M. Olona, M. Rue, and J. Rello, "Impact of diversity of antibiotic use on the development of antimicrobial resistance." *The Journal of antimicrobial chemotherapy*, vol. 57, no. 6, pp. 1197–204, Jun. 2006. [Online]. Available: http://jac.oxfordjournals.org/content/57/6/1197

[254] M. Zasloff, "Antimicrobial peptides of multicellular organisms." *Nature*, vol. 415, no. 6870, pp. 389–395, Jan. 2002. [Online]. Available: http://dx.doi.org/10.1038/415389a

[255] R. E. W. Hancock and H.-G. Sahl, "Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies." *Nat Biotechnol*, vol. 24, no. 12, pp. 1551–1557, Dec. 2006. [Online]. Available: http://dx.doi.org/10.1038/nbt1267

[256] K. V. R. Reddy, R. D. Yedery, and C. Aranha, "Antimicrobial peptides: premises and promises." *International journal of antimicrobial agents*, vol. 24, no. 6, pp. 536–47, Dec. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092485790400322X

[257] A. K. Marr, W. J. Gooderham, and R. E. Hancock, "Antibacterial peptides for therapeutic use: obstacles and realistic outlook." *Current opinion in pharmacology*, vol. 6, no. 5, pp. 468–72, Oct. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1471489206001299

[258] A. Nijnik and R. Hancock, "Host defence peptides: antimicrobial and immunomodulatory activity and potential applications for tackling antibiotic-resistant infections." *Emerging health threats journal*, vol. 2, p. e1, Jan. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3167646&tool=pmcentrez&rendertype=abstract

[259] K. a. Brogden, "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?" *Nature reviews. Microbiology*, vol. 3, no. 3, pp. 238–50, Mar. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15703760

[260] R. E. Hancock, "Host defence (cationic) peptides: what is their future clinical potential?" *Drugs*, vol. 57, no. 4, pp. 469–473, Apr. 1999.

[261] K. A. Brogden, A. J. De Lucca, J. Bland, and S. Elliott, "Isolation of an ovine pulmonary surfactant-associated anionic peptide bactericidal for Pasteurella haemolytica." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 1, pp. 412–6, Jan. 1996. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=40248&tool=pmcentrez&rendertype=abstract

[262] G. Wang, X. Li, and Z. Wang, "APD2: the updated antimicrobial peptide database and its application in peptide design." *Nucleic acids research*, vol. 37, no. Database issue, pp. D933–7, Jan. 2009. [Online]. Available: http://nar.oxfordjournals.org/content/37/suppl_1/D933.abstract

[263] J.-P. S. Powers and R. E. W. Hancock, "The relationship between peptide structure and antibacterial activity." *Peptides*, vol. 24, no. 11, pp. 1681–1691, Nov. 2003. [Online]. Available: http://dx.doi.org/10.1016/j.peptides.2003.08.023

[264] V. Nizet, "Antimicrobial peptide resistance mechanisms of human bacterial pathogens." *Current issues in molecular biology*, vol. 8, no. 1, pp. 11–26, Jan. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16450883

[265] M. W. Bader, S. Sanowar, M. E. Daley, A. R. Schneider, U. Cho, W. Xu, R. E. Klevit, H. Le Moual, and S. I. Miller, "Recognition of antimicrobial peptides by a bacterial sensor kinase." *Cell*, vol. 122, no. 3, pp. 461–72, Aug. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092867405005532

[266] W. M. Shafer, Ed., *Antimicrobial Peptides and Human Disease*, ser. Current Topics in Microbiology and Immunology. Springer Berlin Heidelberg, 2006, vol. 306. [Online]. Available: http://www.springerlink.com/index/10.1007/3-540-29916-5

[267] D. Andreu and L. Rivas, "Animal antimicrobial peptides: an overview." *Biopolymers*, vol. 47, no. 6, pp. 415–433, 1998. [Online]. Available: http://dx.doi.org/3.0.CO;2-D

[268] H. G. Boman, "Peptide antibiotics and their role in innate immunity." *Annu Rev Immunol*, vol. 13, pp. 61–92, 1995. [Online]. Available: http://dx.doi.org/10.1146/annurev.iy.13.040195.000425

[269] G. H. Gudmundsson and B. Agerberth, "Neutrophil antibacterial peptides, multifunctional effector molecules in the mammalian immune system." *J Immunol Methods*, vol. 232, no. 1-2, pp. 45–54, Dec. 1999.

[270] R. E. Hancock and G. Diamond, "The role of cationic antimicrobial peptides in innate host defences." *Trends Microbiol*, vol. 8, no. 9, pp. 402–410, Sep. 2000.

[271] Y. Lai and R. L. Gallo, "AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense." *Trends in immunology*, vol. 30, no. 3, pp. 131–41, Mar. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1471490609000052

[272] R. E. W. Hancock, A. Nijnik, and D. J. Philpott, "Modulating immunity as a therapy for bacterial infections." *Nature reviews. Microbiology*, vol. 10, no. 4, pp. 243–54, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2745

[273] N. Saint, H. Cadiou, Y. Bessin, and G. Molle, "Antibacterial peptide pleurocidin forms ion channels in planar lipid bilayers." *Biochim Biophys Acta*, vol. 1564, no. 2, pp. 359–364, Aug. 2002.

[274] H. Fu, A. Björstad, C. Dahlgren, and J. Bylund, "A bactericidal cecropin-A peptide with a stabilized alpha-helical structure possess an increased killing capacity but no proinflammatory activity." *Inflammation*, vol. 28, no. 6, pp. 337–343, Dec. 2004. [Online]. Available: http://dx.doi.org/10.1007/s10753-004-6644-9

[275] M. E. Houston, L. H. Kondejewski, D. N. Karunaratne, M. Gough, S. Fidai, R. S. Hodges, and R. E. Hancock, "Influence of preformed alpha-helix and alpha-helix induction on the activity of cationic antimicrobial peptides." *J Pept Res*, vol. 52, no. 2, pp. 81–88, Aug. 1998.

[276] M. Uteng, H. v. H. Hauge, P. R. L. Markwick, G. Fimland, D. Mantzilas, J. Nissen-Meyer, and C. Muhle-Goll, "Three-dimensional structure in lipid micelles of the pediocin-like antimicrobial peptide sakacin P and a sakacin P variant that is structurally stabilized by an inserted C-terminal disulfide bridge." *Biochemistry*, vol. 42, no. 39, pp. 11 417–11 426, Oct. 2003. [Online]. Available: http://dx.doi.org/10.1021/bi034572i

[277] M. E. Houston, L. H. Kondejewski, D. N. Karunaratne, M. Gough, S. Fidai, R. S. Hodges, and R. E. Hancock, "Influence of preformed alpha-helix and alpha-helix induction on the activity of cationic antimicrobial peptides." *The journal of peptide research : official journal of the American Peptide Society*, vol. 52, no. 2, pp. 81–8, Aug. 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9727863

[278] H. Leontiadou, A. E. Mark, and S. J. Marrink, "Antimicrobial peptides in action." *J Am Chem Soc*, vol. 128, no. 37, pp. 12 156–12 161, Sep. 2006. [Online]. Available: http://dx.doi.org/10.1021/ja062927q

[279] M. R. Yeaman and N. Y. Yount, "Mechanisms of antimicrobial peptide action and resistance." *Pharmacol Rev*, vol. 55, no. 1, pp. 27–55, Mar. 2003. [Online]. Available: http://dx.doi.org/10.1124/pr.55.1.2

[280] M. Dathe, M. Schümann, T. Wieprecht, A. Winkler, M. Beyermann, E. Krause, K. Matsuzaki, O. Murase, and M. Bienert, "Peptide helicity and membrane surface charge modulate the balance of electrostatic and hydrophobic interactions with lipid bilayers and biological membranes." *Biochemistry*, vol. 35, no. 38, pp. 12 612–12 622, Sep. 1996. [Online]. Available: http://dx.doi.org/10.1021/bi960835f

[281] M. Dathe, T. Wieprecht, H. Nikolenko, L. Handel, W. L. Maloy, D. L. MacDonald, M. Beyermann, and M. Bienert, "Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides." *FEBS Lett*, vol. 403, no. 2, pp. 208–212, Feb. 1997.

[282] K. Matsuzaki, "Why and how are peptidelipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1462, no. 1-2, pp. 1–10, Dec. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005273699001972

[283] M. A. Kol, A. N. C. van Laak, D. T. S. Rijkers, J. A. Killian, A. I. P. M. de Kroon, and B. de Kruijff, "Phospholipid flop induced by transmembrane peptides in model membranes is modulated by lipid composition." *Biochemistry*, vol. 42, no. 1, pp. 231–7, Jan. 2003. [Online]. Available: http://dx.doi.org/10.1021/bi0268403

[284] H. W. Huang, "Molecular mechanism of antimicrobial peptides: the origin of cooperativity." *Biochim Biophys Acta*, vol. 1758, no. 9, pp. 1292–1302, Sep. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.bbamem.2006.02.001

[285] M. N. Melo, R. Ferre, and M. A. R. B. Castanho, "Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations." *Nat Rev Microbiol*, vol. 7, no. 3, pp. 245–250, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2095

[286] M. Bastos, G. Bai, P. Gomes, D. Andreu, E. Goormaghtigh, and M. Prieto, "Energetics and partition of two cecropin-melittin hybrid peptides to model membranes of different composition." *Biophys J*, vol. 94, no. 6, pp. 2128–2141, Mar. 2008. [Online]. Available: http://dx.doi.org/10.1529/biophysj.107.119032

[287] K. Matsuzaki, S. Yoneyama, N. Fujii, K. Miyajima, K. Yamada, Y. Kirino, and K. Anzai, "Membrane permeabilization mechanisms of a cyclic antimicrobial peptide, tachyplesin I, and its linear analog." *Biochemistry*, vol. 36, no. 32, pp. 9799–9806, Aug. 1997. [Online]. Available: http://dx.doi.org/10.1021/bi970588v

[288] L. Xie, X. Ge, H. Tan, L. Xie, Y. Zhang, T. Hart, X. Yang, and P. E. Bourne, "Towards structural systems pharmacology to study complex diseases and personalized medicine." *PLoS computational biology*, vol. 10, no. 5, p. e1003554, May 2014. [Online]. Available: http://dx.plos.org/10.1371/journal.pcbi.1003554

[289] J. D. F. Hale and R. E. W. Hancock, "Alternative mechanisms of action of cationic antimicrobial peptides on bacteria." *Expert Rev Anti Infect Ther*, vol. 5, no. 6, pp. 951–959, Dec. 2007. [Online]. Available: http://dx.doi.org/10.1586/14787210.5.6.951

[290] A. Bhunia, P. N. Domadia, J. Torres, K. J. Hallock, A. Ramamoorthy, and S. Bhattacharjya, "NMR structure of pardaxin, a pore-forming antimicrobial peptide, in lipopolysaccharide micelles: mechanism of outer membrane permeabilization." *J Biol Chem*, vol. 285, no. 6, pp. 3883–3895, Feb. 2010. [Online]. Available: http://dx.doi.org/10.1074/jbc.M109.065672

[291] S. J. Ludtke, K. He, W. T. Heller, T. A. Harroun, L. Yang, and H. W. Huang, "Membrane pores induced by magainin." *Biochemistry*, vol. 35, no. 43, pp. 13 723–13 728, Oct. 1996. [Online]. Available: http://dx.doi.org/10.1021/bi9620621

[292] S. J. Ludtke, K. He, Y. Wu, and H. W. Huang, "Cooperative membrane insertion of magainin correlated with its cytolytic activity." *Biochimica et biophysica acta*, vol. 1190, no. 1, pp. 181–4, Mar. 1994. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8110813

[293] W. L. Maloy and U. P. Kari, "Structure-activity studies on magainins and other host defense peptides." *Biopolymers*, vol. 37, no. 2, pp. 105–22, Jan. 1995. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/7893944

[294] a. J. Mason, W. Moussaoui, T. Abdelrahman, A. Boukhari, P. Bertani, A. Marquette, P. Shooshtarizaheh, G. Moulay, N. Boehm, B. Guerold, R. J. H. Sawers, A. Kichler, M.-H. Metz-Boutigue, E. Candolfi, G. Prévost, and B. Bechinger, "Structural determinants of antimicrobial

and antiplasmodial activity and selectivity in histidine-rich amphipathic cationic peptides." *The Journal of biological chemistry*, vol. 284, no. 1, pp. 119–33, Jan. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18984589

[295] L. S. Vermeer, Y. Lan, V. Abbate, E. Ruh, T. T. Bui, L. J. Wilkinson, T. Kanno, E. Jumagulova, J. Kozlowska, J. Patel, C. A. McIntyre, W. C. Yam, G. Siu, R. A. Atkinson, J. K. W. Lam, S. S. Bansal, A. F. Drake, G. H. Mitchell, and A. J. Mason, "Conformational flexibility determines selectivity and antibacterial, antiplasmodial, and anticancer potency of cationic α-helical peptides." *The Journal of biological chemistry*, vol. 287, no. 41, pp. 34 120–33, Oct. 2012. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3464521&tool=pmcentrez&rendertype=abstract

[296] R. F. Epand, W. L. Maloy, A. Ramamoorthy, and R. M. Epand, "Probing the "charge cluster mechanism" in amphipathic helical cationic antimicrobial peptides." *Biochemistry*, vol. 49, no. 19, pp. 4076–84, May 2010. [Online]. Available: http://dx.doi.org/10.1021/bi100378m

[297] Y. M. Song, S.-T. Yang, S. S. Lim, Y. Kim, K.-S. Hahm, J. I. Kim, and S. Y. Shin, "Effects of l- or d-Pro incorporation into hydrophobic or hydrophilic helix face of amphipathic α-helical model peptide on structure and cell selectivity," *Biochemical and Biophysical Research Communications*, vol. 314, no. 2, pp. 615–621, Feb. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006291X03027621

[298] S.-T. Yang, J. Y. Lee, H.-J. Kim, Y.-J. Eu, S. Y. Shin, K.-S. Hahm, and J. I. Kim, "Contribution of a central proline in model amphipathic alpha-helical peptides to self-association, interaction with phospholipids, and antimicrobial mode of action." *The FEBS journal*, vol. 273, no. 17, pp. 4040–54, Sep. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16889633

[299] L. S. Vermeer, Y. Lan, V. Abbate, E. Ruh, T. T. Bui, L. Wilkinson, T. Kanno, E. Jumagulova, J. Kozlowska, J. Patel, C. A. McIntyre, W. C. Yam, G. Siu, R. A. Atkinson, J. K. W. Lam, S. S. Bansal, A. F. Drake, G. H. Mitchell, and A. J. Mason, "Conformational flexibility determines selectivity and anti-bacterial, -Plasmodium and -cancer potency of cationic α-helical peptides,"

Journal of Biological Chemistry, Aug. 2012. [Online]. Available: http://www.jbc.org/content/early/2012/08/06/jbc.M112.359067http://www.jbc.org/content/early/2012/08/06/jbc.M112.359067.full.pdfhttp://www.jbc.org/content/early/2012/08/06/jbc.M112.359067.full.pdf+html?frame=sidebar

[300] M. Zasloff, "Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor." *Proceedings of the National Academy of Sciences*, vol. 84, no. 15, pp. 5449–5453, Aug. 1987. [Online]. Available: http://www.pnas.org/content/84/15/5449

[301] B. Bechinger and K. Lohner, "Detergent-like actions of linear amphipathic cationic antimicrobial peptides." *Biochim Biophys Acta*, vol. 1758, no. 9, pp. 1529–1539, Sep. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.bbamem.2006.07.001

[302] E. Gazit, A. Boman, H. G. Boman, and Y. Shai, "Interaction of the mammalian antibacterial peptide cecropin P1 with phospholipid vesicles." *Biochemistry*, vol. 34, no. 36, pp. 11 479–11 488, Sep. 1995.

[303] R. E. W. Hancock and A. Rozek, "Role of membranes in the activities of antimicrobial cationic peptides." *FEMS Microbiol Lett*, vol. 206, no. 2, pp. 143–149, Jan. 2002.

[304] C. B. Park, H. S. Kim, and S. C. Kim, "Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions." *Biochemical and biophysical research communications*, vol. 244, no. 1, pp. 253–7, Mar. 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006291X98981591

[305] C. B. Park, K. S. Yi, K. Matsuzaki, M. S. Kim, and S. C. Kim, "Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: the proline hinge is responsible for the cell-penetrating ability of buforin II." *Proc Natl Acad Sci U S A*, vol. 97, no. 15, pp. 8245–8250, Jul. 2000. [Online]. Available: http://dx.doi.org/10.1073/pnas.150518097

[306] A. M. Cole, P. Weis, and G. Diamond, "Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder." *J Biol Chem*, vol. 272, no. 18, pp. 12 008–12 013, May 1997.

[307] K. Yoshida, Y. Mukai, T. Niidome, C. Takashi, Y. Tokunaga, T. Hatakeyama, and H. Aoyagi, "Interaction of pleurocidin and its analogs with phospholipid membrane and their antibacterial activity." *J Pept Res*, vol. 57, no. 2, pp. 119–126, Feb. 2001.

[308] A. Patrzykat, C. L. Friedrich, L. Zhang, V. Mendoza, and R. E. W. Hancock, "Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in Escherichia coli." *Antimicrob Agents Chemother*, vol. 46, no. 3, pp. 605–614, Mar. 2002.

[309] A. J. Mason, I. N. H. Chotimah, P. Bertani, and B. Bechinger, "A spectroscopic study of the membrane interaction of the antimicrobial peptide Pleurocidin." *Mol Membr Biol*, vol. 23, no. 2, pp. 185–194, 2006. [Online]. Available: http://dx.doi.org/10.1080/09687860500485303

[310] A. Yonezawa, J. Kuwahara, N. Fujii, and Y. Sugiura, "Binding of tachyplesin I to DNA revealed by footprinting analysis: significant contribution of secondary structure to DNA binding and implication for biological action," *Biochemistry*, vol. 31, no. 11, pp. 2998–3004, Mar. 1992. [Online]. Available: http://dx.doi.org/10.1021/bi00126a022

[311] H. G. Boman, B. Agerberth, and A. Boman, "Mechanisms of action on Escherichia coli of cecropin P1 and PR-39, two antibacterial peptides from pig intestine." *Infect. Immun.*, vol. 61, no. 7, pp. 2978–2984, Jul. 1993. [Online]. Available: http://iai.asm.org/content/61/7/2978.abstract?ijkey=7acf2d4d98e00f79bf7bd8b98a7620410b2f7160&keytype2=tf_ipsecsha

[312] C. Subbalakshmi and N. Sitaram, "Mechanism of antimicrobial action of indolicidin," *FEMS Microbiology Letters*, vol. 160, no. 1, pp. 91–96, Mar. 1998. [Online]. Available: http://doi.wiley.com/10.1111/j.1574-6968.1998.tb12896.x

[313] A. Almaaytah, S. Tarazi, A. Abu-Alhaijaa, Y. Altall, N. Al-shar'i, K. Bodoor, and Q. Al-Balas, "Enhanced Antimicrobial Activity of AamAP1-Lysine, a Novel Synthetic Peptide Analog Derived from the Scorpion Venom Peptide AamAP1." *Pharmaceuticals (Basel, Switzerland)*, vol. 7, no. 5, pp. 502–16, Jan. 2014. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4035766&tool=pmcentrez&rendertype=abstract

[314] T. Wecke and T. Mascher, "Antibiotic research in the age of omics: from expression profiles to interspecies communication." *The Journal of antimicrobial chemotherapy*, vol. 66, no. 12, pp. 2689–704, Dec. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21930574

[315] M. Pietiäinen, M. Gardemeister, M. Mecklin, S. Leskelä, M. Sarvas, and V. P. Kontinen, "Cationic antimicrobial peptides elicit a complex stress response in Bacillus subtilis that involves ECF-type sigma factors and two-component signal transduction systems." *Microbiology (Reading, England)*, vol. 151, no. Pt 5, pp. 1577–92, May 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15870467

[316] M. Pietiäinen, P. François, H.-L. Hyyryläinen, M. Tangomo, V. Sass, H.-G. Sahl, J. Schrenzel, and V. P. Kontinen, "Transcriptome analysis of the responses of Staphylococcus aureus to antimicrobial peptides and characterization of the roles of vraDE and vraSR in antimicrobial resistance." *BMC genomics*, vol. 10, no. 1, p. 429, Jan. 2009. [Online]. Available: http://www.biomedcentral.com/1471-2164/10/429

[317] J. A. Majchrzykiewicz, O. P. Kuipers, and J. J. E. Bijlsma, "Generic and specific adaptive responses of Streptococcus pneumoniae to challenge with three distinct antimicrobial peptides, bacitracin, LL-37, and nisin." *Antimicrobial agents and chemotherapy*, vol. 54, no. 1, pp. 440–51, Jan. 2010. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2798553&tool=pmcentrez&rendertype=abstract

[318] A. J. Mason, I. N. H. Chotimah, P. Bertani, and B. Bechinger, "A spectroscopic study of the membrane interaction of the antimicrobial peptide Pleurocidin." *Molecular membrane biology*, vol. 23, no. 2, pp. 185–94. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16754361

[319] A. J. Mason, A. Marquette, and B. Bechinger, "Zwitterionic phospholipids and sterols modulate antimicrobial peptide-induced membrane destabilization." *Biophysical journal*, vol. 93, no. 12, pp. 4289–99, Dec. 2007. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2098721&tool=pmcentrez&rendertype=abstract

[320] A. J. Mason, P. Bertani, G. Moulay, A. Marquette, B. Perrone, A. F. Drake, A. Kichler, and B. Bechinger, "Membrane interaction of chrysophsin-1, a histidine-rich antimicrobial peptide from red sea bream."

Biochemistry, vol. 46, no. 51, pp. 15 175–87, Dec. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18052076

[321] J. Oliveros, "VENNY. An interactive tool for comparing lists with Venn diagrams," *BioinfoGP, CNB-CSIC*.

[322] Q. Zheng and X.-J. Wang, "GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis." *Nucleic acids research*, vol. 36, no. Web Server issue, pp. W358–63, Jul. 2008. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2447756&tool=pmcentrez&rendertype=abstract

[323] A. Nocker, M. Caspers, A. Esveld-Amanatidou, J. van der Vossen, F. Schuren, R. Montijn, and R. Kort, "Multiparameter viability assay for stress profiling applied to the food pathogen Listeria monocytogenes F2365." *Applied and environmental microbiology*, vol. 77, no. 18, pp. 6433–40, Sep. 2011. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3187153&tool=pmcentrez&rendertype=abstract

[324] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." *Molecular systems biology*, vol. 2, p. 2006.0008, Jan. 2006. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1681482&tool=pmcentrez&rendertype=abstract

[325] I. Wiegand, K. Hilpert, and R. E. W. Hancock, "Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances." *Nature protocols*, vol. 3, no. 2, pp. 163–75, Jan. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18274517

[326] J. Xia, N. Psychogios, N. Young, and D. S. Wishart, "MetaboAnalyst: a web server for metabolomic data analysis and interpretation." *Nucleic acids research*, vol. 37, no. Web Server issue, pp. W652–60, Jul. 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2703878&tool=pmcentrez&rendertype=abstract

[327] A. J. McBroom and M. J. Kuehn, "Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response." *Mol*

Microbiol, vol. 63, no. 2, pp. 545–558, Jan. 2007. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2958.2006.05522.x

[328] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, "The complete genome sequence of Escherichia coli K-12." *Science (New York, N.Y.)*, vol. 277, no. 5331, pp. 1453–62, Sep. 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9278503

[329] R. a. Welch, V. Burland, G. Plunkett, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S.-R. Liou, a. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. T. Mobley, M. S. Donnenberg, and F. R. Blattner, "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 17 020–4, Dec. 2002. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139262&tool=pmcentrez&rendertype=abstract

[330] M. Novikova, A. Metlitskaya, K. Datsenko, T. Kazakov, A. Kazakov, B. Wanner, and K. Severinov, "The Escherichia coli Yej transporter is required for the uptake of translation inhibitor microcin C." *Journal of bacteriology*, vol. 189, no. 22, pp. 8361–5, Nov. 2007. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2168686&tool=pmcentrez&rendertype=abstract

[331] A. J. Rzepiela, D. Sengupta, N. Goga, and S. J. Marrink, "Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions." *Faraday discussions*, vol. 144, pp. 431–43; discussion 445–81, Jan. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20158042

[332] G. G. Perron, M. Zasloff, and G. Bell, "Experimental evolution of resistance to an antimicrobial peptide." *Proc Biol Sci*, vol. 273, no. 1583, pp. 251–256, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1098/rspb.2005.3301

[333] S. M. Eswarappa, K. K. Panguluri, M. Hensel, and D. Chakravortty, "The yejABEF operon of Salmonella confers resistance to antimicrobial peptides and contributes to its virulence." *Microbiology (Reading,*

England), vol. 154, no. Pt 2, pp. 666–78, Feb. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18227269

[334] E. D. Weinberg, "Iron loading and disease surveillance." *Emerging infectious diseases*, vol. 5, no. 3, pp. 346–52. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2640766&tool=pmcentrez&rendertype=abstract

[335] B. E. Holbein, "Iron-controlled infection with Neisseria meningitidis in mice." *Infection and immunity*, vol. 29, no. 3, pp. 886–91, Sep. 1980. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=551213&tool=pmcentrez&rendertype=abstract

[336] C. M. Litwin and S. B. Calderwood, "Role of iron in regulation of virulence genes." *Clin. Microbiol. Rev.*, vol. 6, no. 2, pp. 137–149, Apr. 1993. [Online]. Available: http://cmr.asm.org/content/6/2/137.abstract

[337] K. M. Papp-Wallace and M. E. Maguire, "Manganese transport and the role of manganese in virulence." *Annual review of microbiology*, vol. 60, pp. 187–209, Jan. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16704341

[338] J.-P. Claverys, "A new family of high-affinity ABC manganese and zinc permeases," *Research in Microbiology*, vol. 152, no. 3-4, pp. 231–243, Apr. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923250801011950

[339] K. Nishino and A. Yamaguchi, "EvgA of the two-component signal transduction system modulates production of the yhiUV multidrug transporter in Escherichia coli." *Journal of bacteriology*, vol. 184, no. 8, pp. 2319–23, Apr. 2002. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=134960&tool=pmcentrez&rendertype=abstract

[340] ——, "Overexpression of the response regulator evgA of the two-component signal transduction system modulates multidrug resistance conferred by multidrug resistance transporters." *Journal of bacteriology*, vol. 183, no. 4, pp. 1455–8, Feb. 2001. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=95021&tool=pmcentrez&rendertype=abstract

[341] A. Kato, H. Ohnishi, K. Yamamoto, E. Furuta, H. Tanabe, and R. Utsumi, "Transcription of emrKY is regulated by the EvgA-EvgS two-component system in Escherichia coli K-12." *Bioscience, biotechnology, and biochemistry*, vol. 64, no. 6, pp. 1203–9, Jun. 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10923791

[342] H. Weber, C. Pesavento, A. Possling, G. Tischendorf, and R. Hengge, "Cyclic-di-GMP-mediated signalling within the sigma network of Escherichia coli." *Mol Microbiol*, vol. 62, no. 4, pp. 1014–1034, Nov. 2006. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2958.2006.05440.x

C

Appendix C: Publications

Metabolomics (2013) 9:1262–1273
DOI 10.1007/s11306-013-0538-5

ORIGINAL ARTICLE

A relationship between *Pseudomonal* growth behaviour and cystic fibrosis patient lung function identified in a metabolomic investigation

Justyna Kozlowska · Damian W. Rivett · Louic S. Vermeer · Mary P. Carroll · Kenneth D. Bruce · A. James Mason · Geraint B. Rogers

Received: 19 March 2013 / Accepted: 12 April 2013 / Published online: 5 May 2013

Abstract Chronic polymicrobial lung infections in adult cystic fibrosis patients are typically dominated by high levels of *Pseudomonas aeruginosa*. Determining the impact of *P. aeruginosa* growth on airway secretion composition is fundamental to understanding both the behaviour of this pathogen in vivo, and its relationship with other potential colonising species. We hypothesised that the marked differences in the phenotypes of clinical isolates would be reflected in the metabolite composition of spent culture media. ^{1}H NMR spectroscopy was used to characterise the impact of *P. aeruginosa* growth on a synthetic medium as part of an in vitro CF lower airways model system. Comparisons of 15 CF clinical isolates were made and four distinct metabolomic clusters identified. Highly significant relationships between *P. aeruginosa* isolate cluster membership and both patient lung function (FEV_1) and spent culture pH were identified. This link between clinical isolate growth behaviour and FEV_1 indicates characterisation of *P. aeruginosa* growth may find application in predicting patient lung function while the significant divergence in metabolite production and consumption observed between CF clinical isolates suggests dominant isolate characteristics have the potential to play both a selective role in microbiota composition and influence pseudomonal behaviour in vivo.

Keywords NMR · Cystic fibrosis · Pseudomonal · Lung function

1 Introduction

Chronic lung disease is the main determinant of morbidity and mortality in cystic fibrosis (CF) (Emerson et al. 2002; Rosenfeld et al. 2001), with bacterial infection considered a key driver in this process (Kosorok et al. 2001). *Pseudomonas aeruginosa* is a species that has long been regarded as a pathogen in the CF lung (CF Foundation 2007) and whose presence is associated with reduced life expectancy (Lyczak et al. 2002). With the exception of end stage disease (Bjarnsholt et al. 2010), *P. aeruginosa,* though common, is only one of many species forming the bacterial microbiota associated with the CF lower airways by adulthood (Rogers et al. 2004; Armougom et al. 2009). Moreover, though the bacterial species reported in CF typically vary markedly between individuals (Stressmann et al. 2011); these colonising species are less phylogenetically diverse than the pool of bacterial species reported as passing transiently through the lower airways of healthy individuals (Rogers et al. 2006). Together, these factors suggest that the development of CF airway bacterial communities is a selective process, and that this selection differs between individuals. Exposure to an "infective dose" equivalence of a given species will be required for "infection" to occur. However, the existence of a group of core species, that is commonly but not universally reported (van der Gast et al. 2011), suggests that infection is not due

Electronic supplementary material The online version of this article (doi:10.1007/s11306-013-0538-5) contains supplementary material, which is available to authorized users.

J. Kozlowska · D. W. Rivett · L. S. Vermeer · K. D. Bruce · A. James Mason · G. B. Rogers (✉)
Institute of Pharmaceutical Science, King's College London, Franklin–Wilkins Building, 150 Stamford Street, London SE1 9NH, UK
e-mail: geraint.b.rogers@gmail.com

M. P. Carroll
Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton SO16 6YD, UK

to chance alone. The factors underpinning selection remain unclear; however the identification of key selective drivers offers the possibility of identifying those patients who are at greatest risk of developing a lower airway infection by a specific pathogen.

There are a number of factors that differ between CF patients that might contribute to such selection. For example, the severity of the underlying impairment of CFTR function (a trans-epithelial ion transport protein) (Dean and Santis 1994), the degree to which oxygen tension in airway secretions is reduced as a result of neutrophilic influx (Kolpen et al. 2010), and antibiotic treatment history (Stressmann et al. 2011; Tunney et al. 2011). Here, we investigate the extent of a further potential selective force; the nutritional characteristics of secretions in the airways. Whilst differences in secretion rheology may arise as a result of the range of CFTR defect severities (Boucher 2004), there is no evidence to suggest that the chemical composition of the secretions differs substantially between individuals at the point when they are produced. However, these secretions are typically colonised by high levels of *P. aeruginosa* (commonly 10^6–10^9 cfu/ml; Aaron et al. 2004; Stressmann et al. 2011). *P. aeruginosa* growth will, in turn, reduce the availability of certain carbon and nitrogen sources, and produce a wide range of metabolites. Such shifts in nutritional sources are known to influence bacterial community composition in other contexts (Resat et al. 2012; Dunaj et al. 2012). Together, these changes will result in an altered growth environment, potentially influencing the likelihood of successful colonisation by new bacterial species entering the lower airways, and the gene expression and growth strategies of *P. aeruginosa* itself (Bernier et al. 2011).

Pseudomonas aeruginosa isolates from different CF patients are known to exhibit a broad range of phenotypic characteristics, employ a number of different growth strategies in vivo, including planktonic and biofilm modes (Ciofu et al. 2012), and exploit a wide range of carbon and nitrogen sources (Frimmersdorf et al. 2010). We therefore hypothesised that *P. aeruginosa* CF airway isolates differ in the manner in which they modify the composition of the airway secretions in which they grow, resulting in significant differences in the nutritional growth environment available to the CF airway bacterial community. Characterisation of such biochemical signatures requires the determination of changes in the levels of a large number of molecules. Consequently, we used ^{1}H NMR spectroscopy to obtain an overview of the compositional changes that occur in a defined synthetic CF medium (SCFM) as a result of the growth of these isolates in an in vitro CF airway model. ^{1}H NMR spectroscopy has been used previously to investigate the growth of *P. aeruginosa* type strain PAO1 in *Luria–Bertani broth*, a standard laboratory medium (Gjersing et al. 2007). Here, we aimed to more closely replicate the physiochemical composition of CF airway secretions in a controlled manner and cultured clinical isolates in SCFM. Previous studies comparing *P. aeruginosa* gene expression in CF sputum with that in similar CF synthetic media have shown bacterial behaviour to be similar in the two contexts (Palmer et al. 2007; Fung et al. 2010). By combining ^{1}H NMR spectroscopy with this CF airway growth model, we were therefore able to assess the degree to which the impact of *P. aeruginosa* growth differed between clinical isolates under conditions approximating those encountered in vivo. We report substantial variation in the observed spent media metabolomes and show that the variation between different clinical isolates is related to variation in clinical measures of respiratory disease.

2 Materials and methods

Sputum samples were collected from 13 adult CF patients with ethical approval from Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26). The collection of these samples has been described previously (van der Gast et al. 2011; Stressmann et al. 2011). *P. aeruginosa* isolates were recovered from these samples by inoculation on *P. aeruginosa* selective medium (CM0559 plus SR0103, Oxoid, Cambridge, UK). A representative colony of the numerically dominant morphotype was selected for each patient. In the case of Patients 12 and 13, two prevalent morphotypes were isolated concurrently, with both carried forward for analysis. Details of isolates, phenotypic characteristics, and the patients from which they were obtained, are presented in Table 1.

Bacterial species diversity in these samples was previously determined by 16S rRNA gene clone sequencing (van der Gast et al. 2011). Spent culture pH was determined at 37 °C using an InLab Micro Pro pH electrode and Mi150 pH meter (Mettler Toledo, Leicester, UK). All *P. aeruginosa* isolates were screened for auxotrophy as described previously (Barth and Pitt, 1995). These data are also presented in Table 1.

2.1 Bacterial growth conditions

Pseudomonas aeruginosa in the CF lung grows in stagnant mucus, an environment that is characterised by microaerophilic and anaerobic conditions (Worlitzsch et al. 2002; Yoon et al. 2002). To reflect these in vivo sputum conditions, the following growth model was employed. A defined synthetic CF medium (SCFM) was used, based on a number of different CF synthetic media described previously (Sriramulu et al. 2004; Palmer et al. 2005; Palmer

Table 1 Information on isolates, the patients that they were obtained from and properties of the sputum sample

Isolate	Metabolomic cluster	Mucoidy	Pigmentation	Auxotrophy	Age	Sex	Genotype II	BMI	Diabetes	Mean FEV1 % pred.	CFPE in 12 months	Diversity	Culture pH	cfu/ml equiv.
CF1	I	Mucoid	Yes	Methionine	30	Male	Unknown	29	No	54.9	3	28	5.32	1.17×10^9 (5.37×10^8)
CF2	IIc	Non-mucoid	No	No	45	Female	Unknown	18.5	Yes	40.2	4	2	6.54	8.69×10^8 (1.26×10^8)
CF3	I	Non-mucoid	No	Proline	47	Male	Unknown	20.7	Yes	33.9	0	5	6.32	2.29×10^{10} (6.09×10^9)
CF4	IIc	Non-mucoid	No	No	30	Female	711 + 3A7G	25	No	38.2	3	4	6.62	4.05×10^9 (9.12×10^8)
CF5	IIc	Non-mucoid	No	No	22	Female	Phe508del	19	No	36.2	5	6	6.61	2.03×10^{10} (6.17×10^9)
CF6	–	Non-mucoid	Yes	No	55	Male	G85E	24.5	No	52.2	2	24	6.26	1.34×10^8 (4.72×10^7)
CF7	IIa	Non-mucoid	Yes	No	21	Female	Phe508del	19	No	56.6	4	2	6.19	1.58×10^9 (9.34×10^8)
CF9	IIb	Non-mucoid	Yes	No	22	Male	Phe508del	17.9	Yes	16.5	7	8	6.93	2.40×10^9 (5.31×10^8)
CF10	I	Non-mucoid	No	No	18	Female	Phe508del	22.5	No	84	3	17	5.49	5.25×10^9 (1.15×10^9)
CF11	IIa	Non-mucoid	Yes	No	24	Female	G542X	23.4	No	72.5	3	5	6.06	7.80×10^9 (2.08×10^9)
CF12	IIb	Mucoid	Yes	No	20	Male	Phe508del	30.4	No	26.8	4	14	6.53	5.25×10^9 (2.13×10^9)
CF12b	I	Mucoid	No	No				30.4	No	72.5	4	14	5.83	1.76×10^9 (5.64×10^8)
CF13	IIa	Mucoid	Yes	No	20	Male	Phe508del	21	No	54.4	4	6	6.65	1.81×10^9 (4.34×10^8)
CF13b	IIb	Mucoid	Yes	No				21	No	26.8	4	14	6.87	8.55×10^9 (3.32×10^9)
CF14	IIc	Non-mucoid	Yes	Methionine	23	Male	Phe508del	20.7	Yes	53.6	4	42	6.58	2.18×10^9 (9.34×10^8)

"Diversity" indicates the number of bacterial species identified by 16S rRNA gene clone sequence analysis, as published previously (Stressmann et al. 2012). CFPE—"cystic fibrosis pulmonary exacerbation". Genotype I was phe508del for all patients. "cfu/ml equiv." refers to mean *P. aeruginosa* cells numbers per ml of spent medium, as determined through Q-PCR enumeration, with standard deviation values given in brackets

et al. 2007; Fung et al. 2010). The SCFM used contained 10 g/L BSA, 10 g/L porcine gastric mucin, 1.4 g/L herring sperm DNA, 10 mM MOPS, 5 g/L egg yolk emulsion, 3.6 μM $FeSO_4$, 51.8 mM NaCl, 2.28 mM NH_4Cl, 2.128 mM L-lysine·HCl, 14.9 mM KCl, 1.78 mM L-alanine, 1.754 mM $CaCl_2$, 1.661 mM L-proline, 1.609 mM L-leucine, 1.549 mM L-glutamate·HCl, 1.446 mM L-serine, 1.3 mM NaH_2PO_4, 1.25 mM Na_2HPO_4, 1.203 mM L-glycine, 1.12 mM L-isoleucine, 1.117 mM L-valine, 1.072 mM L-threonine, 0.827 mM L-aspartate, 0.802 mM L-tyrosine, 0.676 mM L-ornithine·HCl, 0.633 mM L-methionine, 0.606 mM $MgCl_2$, 0.53 mM L-phenylalanine, 0.519 mM L-histidine·HCl, 0.348 mM KNO_3, 0.306 mM L-arginine·HCl, 0.16 mM L-cysteine·HCl, 0.119 mM diethylene triamine pentaacetic acid, 0.013 mM L-tryptophan. The pH of the medium was adjusted to 6.8. Media was sterilised by passage through a a 0.45-μm-pore-size syringe filter, except for porcine gastric mucin, which was sterilised separately by heating at 70 °C for 24 h in 95 % ethyl alcohol, as described previously (Mitsui et al. 1976).

Incubation was performed in 9 ml volumes of SCFM in 15 ml Falcon tubes (BD Biosciences, Oxford, UK) with tight lids, for 72 h at 37 °C, with inversion every 12 h, in order to to replicate reduced oxygen tensions and low relative physical disruption of the CF lower airways. Following incubation, bacterial cells were pelleted by centrifugation at 12,000×*g*, 10 min at 4 °C, with the supernatant transferred to fresh NMR tubes with 10 % v/v D_2O added to provide a deuterium lock signal.

2.2 NMR

^{1}H NMR spectra were recorded on a Bruker Avance 400 MHz spectrometer equipped with a 5 mm QNP probe (Bruker UK Limited, Coventry, UK) with sample isolates tested in triplicate (three independent cultures from the same colony) and kept at room temperature. A zgesgp pulse sequence (Bruker) with excitation sculpting using gradients was used (Hwang and Shaka 1995). The ^{1}H 90 degree pulse was 9.75 μs. For each spectrum, 65,536 data points were acquired with 16 scans. To help in the assignment of the metabolite resonances, J-resolved 2D correlation with pre-saturation during relaxation delay using gradients (J-Res, Bruker) spectra were recorded for some of the samples, using default pulse sequences as provided by Bruker. The spectral width was 20 ppm. Free induction decays were multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. The spectra were Fourier transformed and calibrated to a 2,2,3,3,-D4-3-(Trimethylsilyl) propionic acid sodium salt (TSP-2,2,3,3-D4) reference signal at 0 ppm. Phase correction was performed manually and automatic baseline correction was applied.

2.3 Bacterial quantification

P. aeruginosa density in samples at harvesting was determined by quantitative (Q) PCR enumeration of *opr*L gene copies in total DNA extracts, using a protocol described previously (Feizabadi et al. 2010). All Q-PCR reactions were carried out in a total volume of 25 μl using Taqman® Universal PCR Mastermix (Applied Biosystems, Warrington, UK). Quantitative PCR assays were carried out using the Rotorgene 6000 (Qiagen, Crawley,UK) with a temperature profile of 50 °C for 2 min, 95 °C for 10 min, followed by 45 cycles at 95 °C for 15 s and 60 °C for 60 s. The cycling program was adjusted at 95 °C for 10 min and then 35 cycles of 10 s each at 95 °C (denaturation) followed by 35 s at 60 °C with fluorescent collection (annealing and extension). Analysis was performed in triplicate and the mean reported.

2.4 *P. aeruginosa* strain genotyping

Random Amplified Polymorphic DNA (RAPD) assays were performed for each *P. aeruginosa* isolate as described previously (Renders et al. 1996) using the primer ERIC2 (5′-AAGTAAGTGACTGGGGTGAGCG-3′). In those isolates not distinguishable based on the resulting profile, a further RAPD profile using the primer ERIC1 (5′-ATGTAAGCTCCTGGGGATTCAC-3′) was performed. PCR reactions were performed in 25 μl volumes, containing REDTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich, Dorset, United Kingdom) with the addition of $MgCl_2$ to achieve a final concentration of 2.5 mM. Primers were used at a concentration of 0.5 μM and 50 ng of template DNA. Reactions were performed as follows: An initial denaturation step of 94 °C for 2 min was followed by 32 cycles of denaturation at 94 °C for 1 min, annealing at 25 °C for 1 min, and extension at 72 °C for 2 min, with a final extension step at 72 °C for 10 min. Amplification was carried out by using a GeneAmp PCR System 2400 (Perkin-Elmer), and verified on Tris–acetate-EDTA–agarose gel electrophoresis and analysed using Phoretix 1D advanced software, version 5.0 (Nonlinear Dynamics, Newcastle upon Tyne, UK).

2.5 Multivariate analysis

Pre-processing and orthogonal projection to latent structures discriminant analysis (OPLS-DA) were carried out with software that was developed in our laboratory for a previous study (Vermeer et al. 2012) using the python programming language with numpy and scipy for calculations, and matplotlib for visualization. The nonlinear iterative partial least-squares (NIPALS) algorithm (Andersson 2009) was used for OPLS-DA analysis.

Regions above 9.074 ppm and below 0.116 ppm were excluded because of noise content. The water peak, ethanol and TSP reference signal were also excluded. The spectra were bucketed using 0.02 ppm bin size with additional, manual bucketing applied to adjust for peak shifting as described below, leaving 336 data points per spectrum. Spectra were normalized using probabilistic quotient normalisation (Dieterle et al. 2006).

Principal component analysis (PCA) was used to identify clustering patterns from the major variations between the 49 NMR spectra. For this analysis, spectra were pareto-scaled after normalisation. In order to further investigate the compounds discriminating between clusters and provide a robust statistical analysis of putative cluster membership, each possible cluster was analysed against SCFM using orthogonal partial least-squares discriminant analysis (OPLS-DA). Here, spectra were auto-scaled (variance of every data point normalized to 1). Both normalisation and auto-scaling were included in the cross-validation procedure (Supplementary Table 1). Cross-validation was performed where 75 % of the samples were used as a training set and the remaining 25 % as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class, and that at least one sample from each class was present in the test set. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1 used to choose the number of components, with the additional constraint to use a maximum of 10 components. This double cross-validation was repeated 2,000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This leads to 4 × 2,000 models (in the supplementary information, each of these models leads to a point on the scores plot, but loadings and weights are presented as averages over all these models). Finally, this procedure was repeated with randomly generated class assignments to provide a reference value for Q^2. The chosen number of components minus one was then used as an OPLS filter and a PLS-DA analysis with two components was carried out on the filtered data to yield one predictive and one orthogonal component. Back-scaled loadings (Cloarec et al. 2005) were used to identify resonances with high variance and high weight, therefore the discriminating resonances, and verified against the peak intensity of the original spectra after PQN normalisation.

Peaks that allow the models to distinguish between classes were assigned by comparing chemical shift values and multiplicities from J-resolved NMR spectra to values from the BMRB (Ulrich et al. 2007) and HMDB (Wishart et al. 2009), analysis of published *P. aeruginosa* metabolic data (Son et al. 2007; Frimmersdorf et al. 2010) and NMR spectra generation from individual medium components was used to help in the assignment.

2.6 Relationships between PCA and clinical characteristics

One-way factorial ANOVA were performed to test for significant relationships between the *P. aeruginosa* strain cluster membership and sample characteristics, with a significance threshold of $p < 0.05$. Homogeneity of variance and normality of errors were assessed using the Fligner-Killeen and the Shapiro–Wilk tests respectively prior to the ANOVA. If a factor failed either test a non-parametric Kruskal–Wallis rank sum test was performed. Factors that were found to be significant using the ANOVA were further studied using Tukey's honest significant difference (HSD) as a post hoc test. Correlations between the sample characteristics were performed using Spearman's rho correlations. Statistical analyses were performed using R (v.2.13.0, www.r-project.org).

3 Results

3.1 ^{1}H NMR spectroscopy of *Pseudomonas* CF isolates cultured in an airway model medium

^{1}H NMR spectroscopy has been used previously to investigate the growth of *P. aeruginosa* type strain PAO1 in *Luria–Bertani* broth, a standard laboratory medium (Gjersing et al. 2007). Here, we aimed to more closely replicate the physiochemical composition of CF airway secretions in a controlled manner and cultured clinical isolates in SCFM. Representative 1D ^{1}H NMR spectra are shown for each of the isolate clusters identified by the principal component analysis described below, revealing the effect of culturing either *P. aeruginosa* PAO1 or CF clinical isolates (Fig. 1). Insufficient growth of isolate 6 occurred in SCFM medium and hence it was excluded from further analysis. Spectra generated from either sterile medium, or media inoculated by a particular strain showed a high degree of reproducibility in multiple independent replicates. Although we were successful in identifying metabolite changes that are linked to clinical measures, as described below, adherence to the well validated SCFM growth procedures presented some challenges for the NMR studies and subsequent multivariate analysis. In order to remain relevant to CF airway secretion composition, SCFM contains a number of components, such as mucin and BSA, which could result in peak broadening and loss of resolution due to high viscosity (Supp. Fig. 1). In addition, a number of the strains analysed were highly mucoid and were capable of producing large amounts of exopolysaccharide (EPS). Finally, although less noticeable in the SCFM spectrum, resonances attributable to the, presumably, non-metabolised MOPS buffer dominate the spectra derived from spent media.

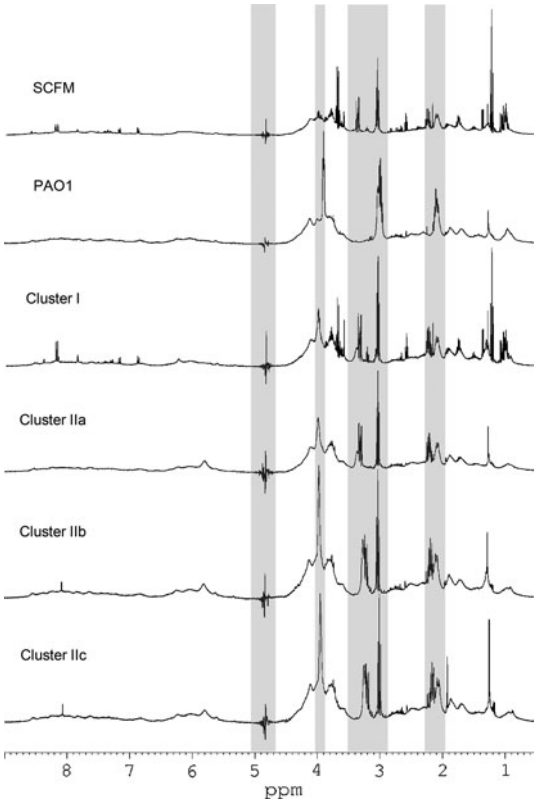


Fig. 1 Representative ^{1}H NMR spectra generated from non-inoculated SCFM, PAO1 inoculated SCFM, and representative members of each of the four putative clinical isolate clusters. *Shaded regions* indicate large regions of the ^{1}H NMR spectra excluded on the basis of solvent or buffer peaks

These factors led to a number of very broad resonances, particularly between 3.5 and 4.5 ppm, and substantial, pH dependent shifting of both broad and sharper resonances between 3.00 and 3.30 ppm and around 2.10 ppm, (Fig. 1) as expected from the pH dependence of buffer resonance chemical shifts (Supp. Fig. 2). To remove the influence of MOPS buffer from the analysis, the following regions were excluded in addition to the water and ethanol peaks (5.02–4.65 ppm; 3.70–3.76 and 1.22–1.15 ppm): 4.02–3.87 ppm; 3.50–2.87 and 2.27–1.96 ppm. Ultimately, this has a number of implications for how the data may be treated and the degree of resonance assignment that is possible. Large scale shifting of buffer resonances resulted in ineffective peak alignment using correlation optimized warping (COW) (Tomasi et al. 2004) (Supp. Fig. 3). Although peak realignment could conceivably be achieved through pH adjustment of spent media as is done for e.g. urine samples (Beneduci et al. 2011), manual peak bucketing was able to account here for the observed peak shifts. The exclusion of large regions, considerable spectral overlap and the appearance of broad resonances following bacterial growth also precluded the use of statistical correlation spectroscopy (STOCSY) or other two-dimensional techniques to aid assignment for many resonances. For this reason our present efforts were largely restricted to using the ^{1}H NMR technique and multivariate analysis to group the isolates according to apparent differences in growth strategies.

3.2 PCA identifies putative clusters based on isolate scores

Principal component analysis (PCA) was used to identify clustering patterns between spectra obtained for the clinical isolates ($n = 41$) and SCFM medium ($n = 8$). The representative 2D scores plots of component 1 (PC1) versus component 2 (PC2), which explain 67.7 % of the variation in the spectra, reveal four putative separate clusters representing the different biochemical composition of the samples as detected by the NMR spectra (Fig. 2). In the PCA scores plot, each data point corresponds to one 1D ^{1}H NMR spectrum, and the reproducibility of the method was supported by the close arrangement of data points corresponding to replicates from each isolate (Fig. 2).

NMR spectra obtained from media inoculated with each of the CF clinical isolates fell into two, readily identifiable, main clusters (I/II) that were distinguished on the basis their relative separation in PC1. The second of these clusters is possibly subdivided into two or three further putative clusters (IIa-c) since the isolates were further separated in PC2, with groups of isolates in separate quadrants of the PCA scores plot. The spectra from isolates in the three or four clusters were each well separated from those of the sterile synthetic media with the exception of isolate 1 which caused almost no change in the ^{1}H NMR spectra of the spent media.

Cluster I was mostly separated from SCFM by PC2. Three of the four isolates found in Cluster I (1, 10, 12b) are notable in that they lead to a very acidic pH in the spent media (Table 1). This might cause the cluster members to be distinguished purely on the basis of pH dropping below pH 6.0 which causes MOPS resonances to shift even beyond the ranges excluded above (Supp. Fig. 2). A broad resonance at 4.04 ppm does appear for these isolates which can be assigned to the MOPS resonance expected in this region, however additional MOPS resonances expected between 3.50 and 3.87 ppm could not be discerned above the contribution from metabolite resonances.

Cluster II was further separated from SCFM by PC1 but subdivision of Cluster II was expected since isolates were also well distributed along PC2 and were separated into upper and lower quadrants. Initially a three cluster model was considered with Cluster IIa in the lower quadrant and Cluster IIb in the upper quadrant. The existence of a fourth putative Cluster IIc, with isolates located intermediate to Clusters IIa and IIb, was considered and tested by OPLS-DA below. Key resonances whose variation contributes to PC1

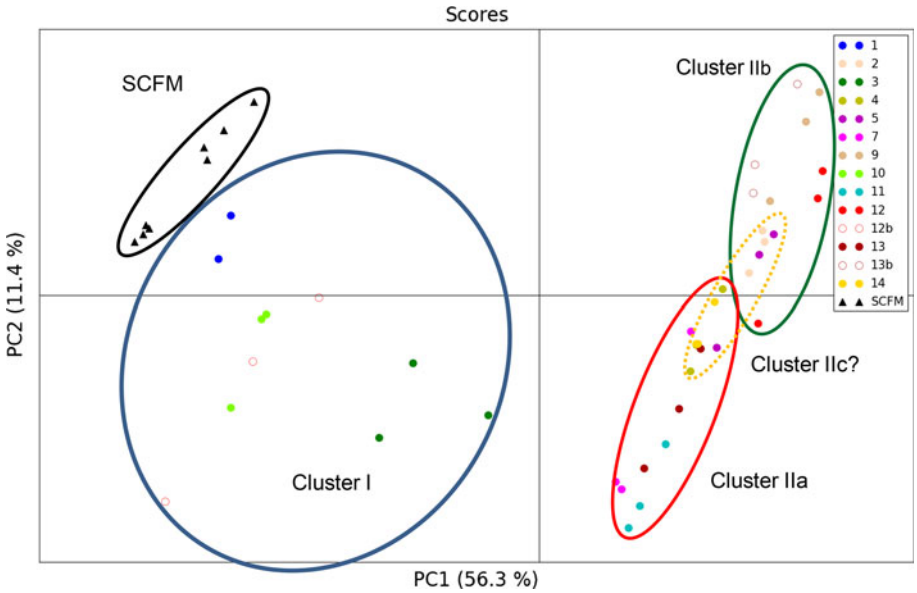


Fig. 2 Scores scatter plot resulting from applying PCA to the ^{1}H NMR data by component 1(PC1) and component 2 (PC2). The percentage of variance in the data explained by each component is indicated on the relevant axis. Strain identification numbers are shown. *Ellipses* are drawn to show putative clusters of spectra. SCFM—synthetic cystic fibrosis media. The corresponding loadings plot provided in the supplementary material identifies points in the NMR spectra that align with either PC1 or PC2

Table 2 Predictive Q^2 values for all models

Test	3 cluster model	4 cluster model
SCFM versus cluster I	0.92 (−0.63)	0.92 (−0.63)
SCFM versus cluster IIa	0.99 (−0.44)	0.99 (−0.47)
SCFM versus cluster IIb	0.99 (−0.46)	0.99 (−0.48)
SCFM versus cluster IIc	–	0.99 (−0.40)
Cluster IIa versus cluster IIb	0.71 (−0.46)	0.91 (−0.50)
Cluster IIa versus cluster IIc	–	0.84 (−0.47)
Cluster IIb versus cluster IIc	–	0.92 (−0.44)

Cluster IIa and IIb contain isolates 4, 7, 11, 13, 14 and 2, 5, 9, 12, 13b, respectively, in the 3 cluster model and lose isolates 2, 4, 5, 14 to Cluster IIc in the four cluster model. Q^2 values for models run with permutated class assignments are given in parentheses

and PC2 are shown in the corresponding PC loadings plot (Supp. Fig. 4).

3.3 OPLS-DA supports clusters identification

Orthogonal projection to latent structures discriminant analysis (OPLS-DA) was then used to compare ^{1}H NMR spectra data generated from sterile synthetic media with each of the four putative strain clusters and to test whether the clusters could indeed be considered separate. Cross-validation was performed on all models (Supp. Fig. 5–8). The resulting 2D scores plots show good separation between the three or four putative clusters and SCFM with Q^2 values >0.90, indicating a highly reliable model compared with an ideal score of 1 (Table 2). Initially, the PCA scores plot readily identified two separate isolate scores clusters (Cluster I/Cluster II) with Cluster I, comprising isolates 1, 3, 10 and 12b, well removed from the remaining isolates. The subdivision of Cluster II was tested; putative Clusters IIa and IIb, containing isolates 4, 7, 11, 13, 14 and 2, 5, 9, 12, 13b respectively, could be separated (Supp. Fig. 8; $Q^2 = 0.69$) using OPLS-DA. However when a further putative cluster (IIc) was considered, comprising isolates 2, 4, 5 and 14 (PCA scores for these isolates show an intermediate distribution between the upper and lower quadrants due to PC2), the apparent separation as monitored by scores plots (Suppl. Fig. 6/7) and Q^2 (Table 2) indicated that a four cluster model may be useful when greater numbers of patient isolates are available.

3.4 Relationships between strain cluster membership and sample characteristics with 1D ^{1}H NMR spectra

To determine whether differences in the nutritional modifications to airway secretion composition that result from the growth of *P. aeruginosa* may have clinical impacts, membership of CF sputum isolate clusters, as defined based on PCA and OPLS-DA, for both three and four cluster models, was compared with a number of potentially key strain or sputum sample characteristics. These factors were isolate auxotrophy, mucoidy, pigmentation, spent culture pH, sputum pH, species richness and relative phyla abundance as defined by 16S rRNA gene clone sequencing analysis, *P. aeruginosa* density as determined by quantitative PCR, patient age, sex, genotype, BMI, diabetic status, FEV_1, and the number of respiratory exacerbations over the preceding 12 months. Of these, highly significant relationships were identified between both cluster membership and lung function (FEV_1) ($F(3,10) = 5.64$, $p = 0.0159$) and cluster membership and spent culture pH ($F(3,10) = 8.63$, $p = 0.004$,) (Fig. 3). These significant relationships were tested using Tukey's HSD to assess for significant differences between clusters. In the three cluster model, this analysis found significant differences between

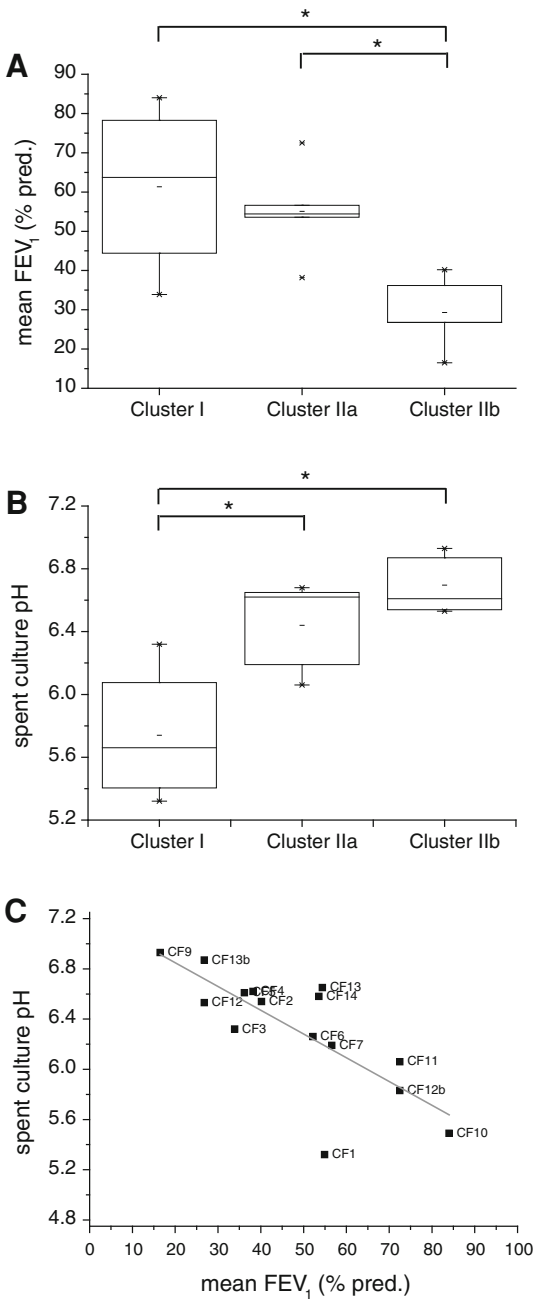


Fig. 3 *Box plots* comparing FEV_1 (**a**) and spent culture pH (**b**) for each of the clusters in the three cluster model; $*p < 0.05$. The relationship between mean FEV_1 and spent culture pH is also shown for each of the isolates (**c**)

Clusters I and IIb ($p_{adj} = 0.020$) and between Clusters IIa and IIb ($p_{adj} = 0.030$) for lung function (FEV_1) with patients in Cluster IIb having relatively poor lung function (Fig. 3a). In the four cluster model, Cluster I was shown to be significantly different from both clusters IIb ($p_{adj} = 0.005$) and IIc ($p_{adj} = 0.010$) (Supp. Fig. 9). A possible relationship was also observed between sputum pH and cluster membership ($F(3,10) = 3.06$, $p = 0.078$, pH ranged from 5.9 to 7.8). A strong negative correlation was found between FEV_1 and spent culture pH ($R = -0.76$, $p = 0.002$) (Fig. 3c) and significant differences in spent culture pH were observed between Cluster I and both Clusters IIa and IIb in the three cluster model (Fig. 3b). No significant correlation was found however between lung function and sputum pH ($R = 0.50$, $p = 0.067$) or between sputum pH and spent culture pH ($R = -0.37$, $p = 0.188$) (further significant correlations are shown in supplementary information as are box plots for the four cluster model; Supp. Fig. 10). Therefore, although Cluster I and Cluster II are clearly separated in the PCA analysis (by PC1), the only significant differences that were found both in FEV_1 and spent culture pH were between Cluster I and Cluster IIb with a significant difference in FEV_1 also seen between Clusters IIa and IIb. These clusters are separated in the PCA analysis by PC2 and hence identification of resonances contributing to PC2, rather than PC1, or OPLS-DA analyses between these clusters should identify metabolomic changes that correlate with variance in FEV_1 or spent culture pH.

Pseudomonas aeruginosa cell numbers in spent media were determined by Q-PCR enumeration, with mean values calculated from analysis of triplicate independent repeat cultures (Table 1). No significant difference between cultures of separate isolates, or independent repeat cultures of particular isolates, was observed, and no relationship was found between *P. aeruginosa* levels and cluster membership.

To determine whether the *P. aeruginosa* isolates belonging to separate clusters were genetically distinct, or represented the same strains growing differently, RAPD PCR analysis was performed. All 15 isolates studied here were found to represent distinct strain types (Supp. Fig. 11).

3.5 OPLS-DA identifies characteristic metabolite consumption and production

The OPLS-DA comparisons of Cluster IIb with Cluster I (Fig. 4a) and Cluster IIa (Fig. 4b) again support the clustering determined above and allow identification of resonances from metabolites that may be implicated in FEV_1 and/or spent culture pH. Plotting normalised spectra from each of the four clusters, coloured according to cluster, highlights resonances whose intensity is consistently altered between clusters (Fig. 5). Comparing the back-scaled loadings plots for the comparisons between Cluster IIb and Cluster I and between Cluster IIb and Cluster IIa identifies metabolites whose differing intensities correlate with the significant differences identified above for lung function FEV_1 and/or spent culture pH. Notably, levels of lysine or ornithine appear higher in the spent media of

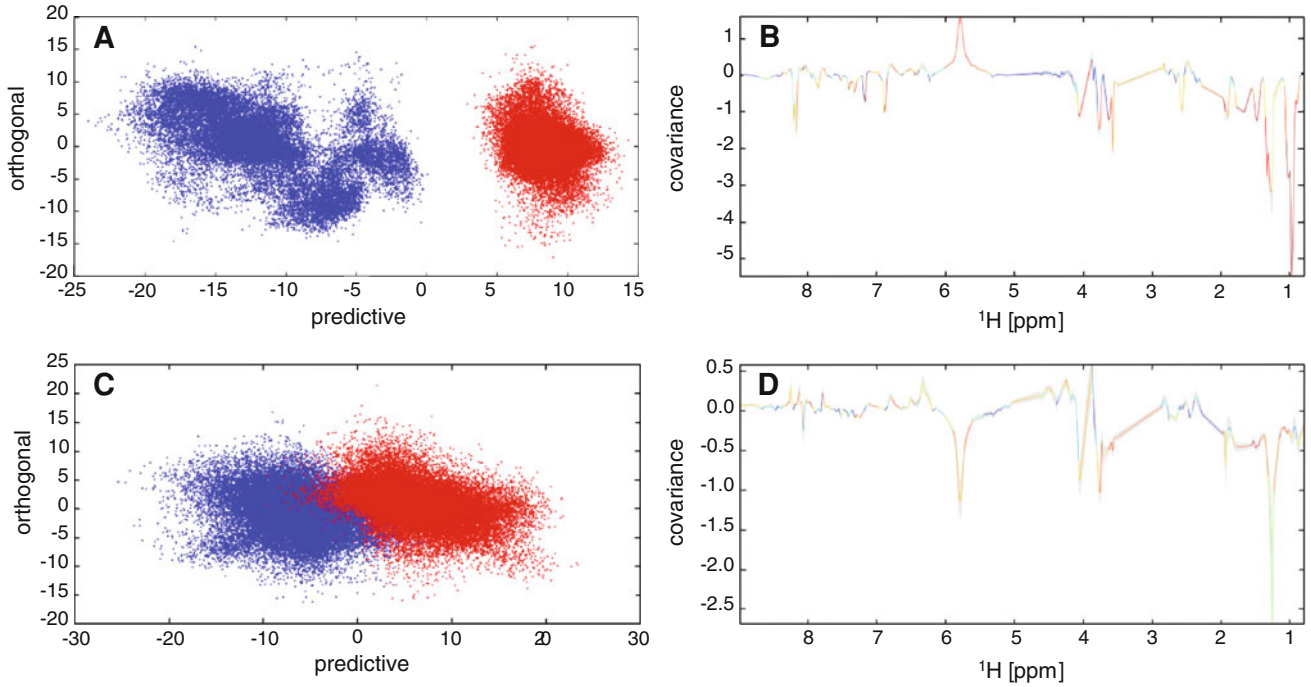


Fig. 4 OPLS-DA scores plots (**a**, **c**) and back-scaled loadings plots (**b**, **d**) for the comparisons between Cluster IIb and Cluster I (**a**, **b**) and between Cluster IIb and IIa (**c**, **d**). Complete Leave-one-out cross-validation (LOOCV) output files are provided in the Supplementary Material

isolates in Cluster I or IIa when compared with those from Cluster IIb, where lung function was poorest, as evidenced by their characteristic resonances at 3.765 ppm (Fig. 5a) and 1.465 ppm (Fig. 5b). A resonance characteristic of leucine at 1.728 ppm (Fig. 5b) is also elevated in spectra of spent media of isolates in both Clusters I and IIa while a broad resonance that appears at 6.80 ppm (Fig. 5c), in many spectra from Cluster IIb isolates, is largely absent from either Clusters I or IIa. Spectra from isolates in Cluster IIc are intermediate between spectra from isolates in either Cluster IIa or IIb for these features. An additional broad resonance at 5.77 ppm is notable (Fig. 5c) but its intensity does not correlate with FEV_1 and is one of the main resonances that contributes to the separation of the isolates by PC1 in the PCA analysis.

4 Discussion

The way in which chronic colonisation by high cell numbers of *P. aeruginosa* affects the composition of airway secretions in the CF lung is likely to be important in selecting co-infecting bacterial species, and in modifying the growth of *P. aeruginosa* itself, since nutritional sources are known to influence bacterial community composition in other contexts (Resat et al. 2012; Dunaj et al. 2012).

Bacterial growth results in both large- and fine-scale modifications to the composition of the growth medium, with the former primarily occurring through the utilisation of compounds as carbon and nitrogen sources and the production of major metabolites. Since the aim here was to obtain an indication of the degree to which *P. aeruginosa* isolate growth might have a differential impact on CF airway secretions, an approach that provides a comprehensive overview of major compositional changes, ^{1}H NMR spectroscopy, was employed (Gjersing et al. 2007). In keeping with previous studies using ^{1}H NMR spectroscopy, the data presented here show a high degree of independent-replicate reproducibility, whilst allowing differentiation of chemically divergent samples.

A defined synthetic growth medium designed to replicate the composition of CF lower airway secretions was used as part of an in vitro model of CF airway conditions. Previous studies comparing *P. aeruginosa* gene expression in CF sputum with that in similar CF synthetic media have shown bacterial behaviour to be similar in the two contexts (Palmer et al. 2007; Fung et al. 2010). By combining ^{1}H NMR spectroscopy with this CF airway growth model, we were therefore able to assess the degree to which the impact of *P. aeruginosa* growth differed between clinical isolates under conditions approximating those encountered in vivo.

The *P. aeruginosa* clinical isolates studied here showed substantial metabolomic differences, and were categorised into four separate clusters. These variations, particularly in levels of the amino acids, may reflect differences in energy strategies employed, and the ability of strains to adapt to

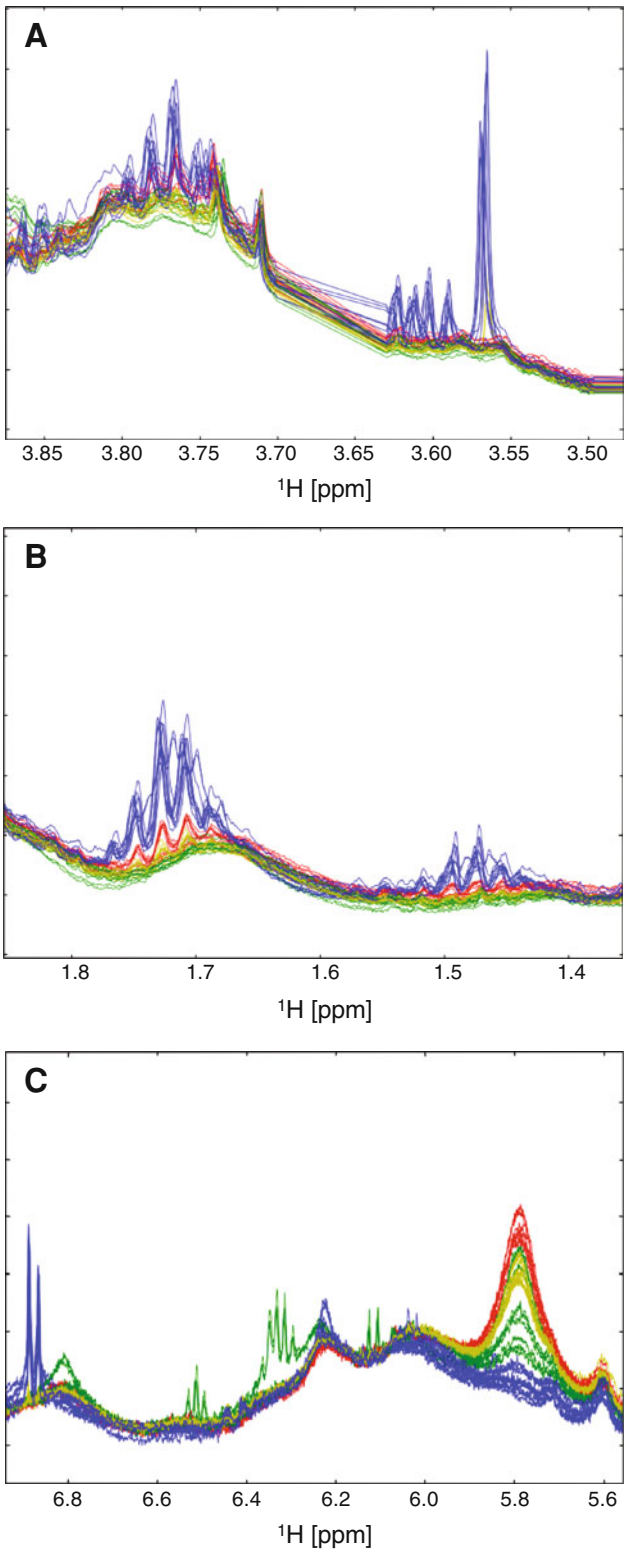


Fig. 5 Normalised ^{1}H spectra (with excluded regions but otherwise untreated) of spent media coloured according to cluster membership. (Cluster I—*blue*, Cluster IIa—*red*, Cluster IIb—*green*, Cluster IIc—*yellow*). Spectral regions between 3.5 and 3.85 ppm (**a**), 5.6 and 6.9 ppm (**b**) and 1.4 and 1.8 ppm (**c**) are shown (Color figure online)

the low oxygen tensions in muco-purulent CF sputum. Such differences may be important in vivo given that amino acid levels have been shown to influence the antibiotic susceptibility of *P. aeruginosa* (Nguyen et al. 2011) and the growth strategies that they adopt (Bernier et al. 2011; Shrout and Chopp 2006).

Cluster membership was found to be related to culture pH. Care was taken to exclude the contribution of pH induced peak shifting through the exclusion of resonances from MOPS buffer prior to multivariate analysis and manual bucketing of spectra following the failure of peak alignment using COW. Visual inspection of the spectra resulting from each cluster suggested that the observed isolate clustering is due to differences in levels of media constituents and metabolites produced/consumed rather than shifting peaks. Indeed, no relationship was found between cluster membership and *P. aeruginosa* cell numbers, further supporting divergence driven by growth strategy rather than cell density. The mechanism by which *P. aeruginosa* growth affects pH is not clear, and could occur either through a change in levels of non-pH neutral components of the growth media or through proton extrusion. However, the impact of *P. aeruginosa* growth on pH could have major clinical implications since it has been shown to influence both bacterial community composition (Romanowski et al. 2011; Duncan et al. 2009) and behaviour (Walker and Duncan 2005) in other clinical contexts, and alterations in airway secretion pH could affect a number of innate defence processes, such as ciliary function (Clary-Meinesz et al. 1998) and mucus viscosity (Inglis et al. 1998).

Importantly, a highly significant relationship was found between the cluster membership of CF isolates analysed here and the lung function of the patient from which they were obtained (as measured by FEV_1). This link between the nutritional modification of the environment by *P. aeruginosa* growth and patient lung function has clear implications for how our understanding of bacterial community composition in the CF airways needs to develop. A broad phylogenetic range of species have been reported in CF respiratory secretion (Rogers et al. 2004), with a common factor for their establishment in the airways being access to nutrients they require for growth. However, not all species common in associated areas, such as the oral cavity, are routinely reported in CF lower airway secretions. This may be influenced by modification of the nutritional context of the lower airway secretions by dominant bacterial species (Kloosterman and Kuipers 2011; Rogosa and Bishop 1964; Nakada and Itoh 2003). Such nutritional interactions in the CF lower airways are poorly understood, and require further investigation.

5 Conclusions

In summary, the application of ^{1}H NMR here to determining the impact of *P. aeruginosa* clinical isolate growth within a model CF system reveals substantial metabolomic differences between isolates. Membership of isolate clusters, defined through PCA, appears to be linked to divergence in metabolite production, with a significant correlation between cluster membership and spent culture pH. These findings suggest that *P. aeruginosa* isolates employ a range of growth strategies. Further, cluster membership was found to be significantly correlated with patient lung function, suggesting that there may be direct clinical implications for bacterial metabolic strategy in vivo. A more sophisticated characterisation of the metabolic and pH environment of the CF lower airway is now warranted, with the potential to inform our understanding of the chronic bacterial infections that typify cystic fibrosis airway disease.

Acknowledgments This study was supported by the Anna Trust. JK is supported by a BBSRC Industrial CASE studentship.

References

Aaron, S. D., Kottachchi, D., Ferris, W. J., Vandemheen, K. L., St Denis, M. L., Plouffe, A., et al. (2004). Sputum versus bronchoscopy for diagnosis of *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *European Respiratory Journal, 24*, 631–637.

Andersson, M. (2009). A comparison of nine PLS1 algorithms. *Journal of Chemometrics, 23*, 518–529.

Armougom, F., Bittar, F., Stremler, N., Rolain, J. M., Robert, C., Dubus, J. C., et al. (2009). Microbial diversity in the sputum of a cystic fibrosis patient studied with 16S rDNA pyrosequencing. *European Journal of Clinical Microbiology and Infectious Diseases, 28*, 1151–1154.

Barth, A. L., & Pitt, T. L. (1995). Auxotrophy of *Burkholderia* (*Pseudomonas*) *cepacia* from cystic fibrosis patients. *Journal of Clinical Microbiology, 33*, 2192–2194.

Beneduci, A., Chidichimo, G., Dardo, G., & Pontoni, G. (2011). Highly routinely reproducible alignment of ^{1}H NMR spectral peaks if metabolites in huge sets of urines. *Analytica Chimica Acta, 685*, 186–195.

Bernier, S. P., Ha, D. G., Khan, W., Merritt, J. H., & O'Toole, G. A. (2011). Modulation of *Pseudomonas aeruginosa* surface-associated group behaviors by individual amino acids through c-di-GMP signaling. *Research in Microbiology, 162*, 680–688.

Bjarnsholt, T., Jensen, P. Ø., Jakobsen, T. H., Phipps, R., Nielsen, A. K., Rybtke, M. T., et al. (2010). Scandinavian cystic fibrosis study consortium. Quorum sensing and virulence of *Pseudomonas aeruginosa* during lung infection of cystic fibrosis patients. *PLoS ONE, 5*, e10115.

Boucher, R. C. (2004). New concepts of the pathogenesis of cystic fibrosis lung disease. *European Respiratory Journal, 23*, 146–158.

CF Foundation. 2007. Patient registry annual data report. http://cff.org/UploadedFiles/research/ClinicalResearch/2007-Patient-Registry-Report.pdf.

Ciofu, O., Mandsberg, L. F., Wang, H., & Høiby, N. (2012). Phenotypes selected during chronic lung infection in cystic fibrosis patients: implications for the treatment of *Pseudomonas aeruginosa* biofilm infections. *FEMS Immunology and Medical Microbiology, 65*, 215–225.

Clary-Meinesz, C., Mouroux, J., Cosson, J., Huitorel, P., & Blaive, B. (1998). Influence of external pH on ciliary beat frequency in human bronchi and bronchioles. *European Respiratory Journal, 11*, 330–333.

Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., et al. (2005). Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^{1}H NMR spectroscopic metabonomic studies. *Analytical Chemistry, 77*, 517–526.

Dean, M., & Santis, G. (1994). Heterogeneity in the severity of cystic fibrosis and the role of CFTR gene mutations. *Human Genetics, 93*, 364–368.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalisation as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Analytical Chemistry, 78*, 4281–4290.

Dunaj, S. J., Vallino, J. J., Hines, M. E., Gay, M., Kobyljanec, C., & Rooney-Varga, J. N. (2012). Relationships between soil organic matter, nutrients, bacterial community structure, and the performance of microbial fuel cells. *Environmental Science and Technology, 46*, 1914–1922.

Duncan, S. H., Louis, P., Thomson, J. M., & Flint, H. J. (2009). The role of pH in determining the species composition of the human colonic microbiota. *Environmental Microbiology, 11*, 2112–2122.

Emerson, J., Rosenfeld, M., McNamara, S., Ramsey, B., & Gibson, R. L. (2002). *Pseudomonas aeruginosa* and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatric Pulmonology, 34*, 91–100.

Feizabadi, M. M., Majnooni, A., Nomanpour, B., Fatolahzadeh, B., Raji, N., Delfani, S., et al. (2010). Direct detection of *Pseudomonas aeruginosa* from patients with healthcare associated pneumonia by real time PCR. *Infection, Genetics and Evolution, 10*, 1247–1251.

Frimmersdorf, E., Horatzek, S., Pelnikevich, A., Wiehlmann, L., & Schomburg, D. (2010). How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach. *Environmental Microbiology, 12*, 1734–1747.

Fung, C., Naughton, S., Turnbull, L., Tingpej, P., Rose, B., Arthur, J., et al. (2010). Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. *Journal of Medical Microbiology, 59*, 1089–1100.

Gjersing, E. L., Herberg, J. L., Horn, J., Schaldach, C. M., & Maxwell, R. S. (2007). NMR metabolomics of planktonic and biofilm modes of growth in *Pseudomonas aeruginosa*. *Analytical Chemistry, 79*, 8037–8045.

Hwang, T.-L., & Shaka, A. J. (1995). Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, 112*, 275–279.

Inglis, S. K., Corboz, M. R., & Ballard, S. T. (1998). Effect of anion secretion inhibitors on mucin content of airway submucosal gland ducts. *American Journal of Physics, 274*, L762–L766.

Kloosterman, T. G., & Kuipers, O. P. (2011). Regulation of arginine acquisition and virulence gene expression in the human pathogen *Streptococcus pneumoniae* by transcription regulators ArgR1 and AhrC. *Journal of Biological Chemistry, 286*, 44594–44605.

Kolpen, M., Hansen, C. R., Bjarnsholt, T., Moser, C., Christensen, L. D., van Gennip, M., et al. (2010). Polymorphonuclear leucocytes consume oxygen in sputum from chronic *Pseudomonas aeruginosa* pneumonia in cystic fibrosis. *Thorax, 65*, 57–62.

Kosorok, M. R., Zeng, L., West, S. E., Rock, M. J., Splaingard, M. L., Laxova, A., et al. (2001). Acceleration of lung disease in children with cystic fibrosis after *Pseudomonas aeruginosa* acquisition. *Pediatric Pulmonology, 32*, 277–287.

Lyczak, J. B., Cannon, C. L., & Pier, G. B. (2002). Lung infections associated with cystic fibrosis. *Clinical Microbiology Reviews, 15*, 194–222.

Mitsui, Y., Matsumura, K., Kondo, C., & Takashima, R. (1976). The role of mucin on experimental *Pseudomonas keratitis* in rabbits. *Investigative Ophthalmology, 15*, 208–210.

Nakada, Y., & Itoh, Y. (2003). Identification of the putrescine biosynthetic genes in *Pseudomonas aeruginosa* and characterization of agmatine deiminase and N-carbamoylputrescine amidohydrolase of the arginine decarboxylase pathway. *Microbiology, 149*, 707–714.

Nguyen, D., Joshi-Datar, A., Lepine, F., Bauerle, E., Olakanmi, O., Beer, K., et al. (2011). Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science, 334*, 982.

Palmer, K. L., Aye, L. M., & Whiteley, M. (2007). Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *Journal of Bacteriology, 189*, 8079–8087.

Palmer, K. L., Mashburn, L. M., Singh, P. K., & Whiteley, M. (2005). Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. *Journal of Bacteriology, 187*, 5267–5277.

Renders, N., Römling, Y., Verbrugh, H., & van Belkum, A. (1996). Comparative typing of *Pseudomonas aeruginosa* by random amplification of polymorphic DNA or pulsed-field gel electrophoresis of DNA macrorestriction fragments. *Journal of Clinical Microbiology, 34*, 3190–3195.

Resat, H., Bailey, V., McCue, L. A., & Konopka, A. (2012). Modeling microbial dynamics in heterogeneous environments: growth on soil carbon sources. *Microbial Ecology, 63*, 883–897.

Rogers, G. B., Carroll, M. P., Serisier, D. J., Hockey, P. M., Jones, G., & Bruce, K. D. (2004). Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *Journal of Clinical Microbiology, 42*, 5176–5183.

Rogers, G. B., Carroll, M. P., Serisier, D. J., Hockey, P. M., Jones, G., Kehagia, V., et al. (2006). Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. *Journal of Clinical Microbiology, 44*, 2601–2604.

Rogosa, M., & Bishop, F. S. (1964). The genus *Veillonella* II. Nutritional studies. *Journal of Bacteriology, 87*, 574–580.

Romanowski, K., Zaborin, A., Fernandez, H., Poroyko, V., Valuckaite, V., Gerdes, S., et al. (2011). Prevention of siderophore-mediated gut-derived sepsis due to *P. aeruginosa* can be achieved without iron provision by maintaining local phosphate abundance: role of pH. *BMC Microbiology, 11*, 212.

Rosenfeld, M., Emerson, J., Williams-Warren, J., Pepe, M., Smith, A., Montgomery, A. B., et al. (2001). Defining a pulmonary exacerbation in cystic fibrosis. *Journal of Pediatrics, 139*, 359–365.

Shrout, J. D., Chopp, D. L., Just, C. L., Hentzer, M., Givskov, M., & Parsek, M. R. (2006). The impact of quorum sensing and swarming motility on Pseudomonas aeruginosa biofilm formation is nutritionally conditional. *Molecular Microbiology, 62*, 1264–1277.

Son, M. S., Matthews, W. J, Jr, Kang, Y., Nguyen, D. T., & Hoang, T. T. (2007). *In vivo* evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infection and Immunity, 75*, 5313–5324.

Sriramulu, D. D., Lünsdorf, H., Lam, J. S., & Römling, U. (2004). Microcolony formation: a novel biofilm model of *Pseudomonas aeruginosa* for the cystic fibrosis lung. *Journal of Medical Microbiology, 54*, 667–676.

Stressmann, F. A., Rogers, G. B., Marsh, P., Lilley, A. K., Daniels, T. W., Carroll, M. P., et al. (2011). Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of Cystic Fibrosis, 10*, 357–365.

Stressmann, F. A., Rogers, G. B., van der Gast, C. J., Marsh, P., Vermeer, L. S., Carroll, M. P., et al. (2012). Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax, 67*, 867–873.

Tomasi, G., van den Berg, F., & Andersson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics, 18*, 231–241.

Tunney, M. M., Klem, E. R., Fodor, A. A., Gilpin, D. F., Moriarty, T. F., McGrath, S. J., et al. (2011). Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax, 66*, 579–584.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., et al. (2007). BioMagResBank. *Nucleic Acids Research, 36*, D402–D408.

van der Gast, C. J., Walker, A. W., Stressmann, F. A., Rogers, G. B., Scott, P., Daniels, T. W., et al. (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME Journal, 5*, 780–791.

Vermeer, L. S., Fruhwirth, G. O., Pandya, P., Ng, T., & Mason, A. J. (2012). NMR metabolomics of MTLn3E breast cancer cells identifies a role for CxCR4 in lipid and choline regulation. *Journal of Proteome Research, 11*, 2996–3003.

Walker, A. W., & Duncan, S. H. (2005). McWilliam Leitch EC, Child MW, Flint HJ. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Applied and Environmental Microbiology, 71*, 3692–3700.

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., et al. (2009). HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research, 37*, D603–D610.

Worlitzsch, D., Tarran, R., Ulrich, M., Schwab, U., Cekici, A., Meyer, K. C., et al. (2002). Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *The Journal of Clinical Investigation, 109*, 317–325.

Yoon, S. S., Hennigan, R. F., Hilliard, G. M., Ochsner, U. A., Parvatiyar, K., Kamani, M. C., et al. (2002). *Pseudomonas aeruginosa* anaerobic respiration in biofilms. Relationships to cystic fibrosis pathogenesis. *Developmental Cell, 3*, 593–603.

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MOLECULAR BIOLOGY
MICROBIOLOGY

Received
17 March 2014

Accepted
5 June 2014

Published
25 June 2014

Correspondence and requests for materials should be addressed to G.B.R. (geraint.b.rogers@gmail.com)

Functional divergence in gastrointestinal microbiota in physically-separated genetically identical mice

G. B. Rogers[1,2,3], J. Kozlowska[2], J. Keeble[2], K. Metcalfe[4], M. Fao[4], S. E. Dowd[5], A. J. Mason[2], M. A. McGuckin[1] & K. D. Bruce[2]

[1]Immunity, Infection, and Inflammation Program, Mater Research Institute – University of Queensland, Translational Research Institute, Woolloongabba, Australia, [2]King's College London, Institute of Pharmaceutical Science, London, SE1 9NH, UK, [3]SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Bedford Park, Adelaide, Australia, [4]Charles River UK, Manston Rd. Margate, Kent CT9 4LT UK, [5]Molecular Research MR DNA, Shallowater, TX 79363, USA.

Despite the fundamental contribution of the gut microbiota to host physiology, the extent of its variation in genetically-identical animals used in research is not known. We report significant divergence in both the composition and metabolism of gut microbiota in genetically-identical adult C57BL/6 mice housed in separate controlled units within a single commercial production facility. The reported divergence in gut microbiota has the potential to confound experimental studies using mammalian models.

Researchers using animal models are becoming increasingly aware of possible influences of the gut microbiota on physiology. Murine models have been used to demonstrate relationships between the gut microbiota and obesity[1], metabolic disease[2], cardiovascular health[3], nervous system development[4], diabetes[5], and immune function[6], hepatic function[7], inflammatory bowel conditions[8], and carcinogenesis[9], highlighting the potential impact that differences in the microbiome of mice from different animal facilities could have on research. However, most researchers assume that genetically-identical mice derived from a single supplier will have an equivalent microbiome. To test this assumption we studied the faecal microbiome and metabolome of genetically-identical C57BL/6 mice housed in four separate controlled units within a single facility of a commercial supplier of animals for research. Faecal samples were collected at eight weeks of age from twenty mice, with five mice sampled in each of four barrier rooms. These mice were separated by no more than ten generations.

Methods

Murine faecal samples. Faeces were collected from eight week old C57BL/6 at the Charles River commercial facility (Margate, UK) under commercial licence, with all mice kept in accordance with protocols approved by The Animal Health and Welfare Board for England. Samples were collected from 20 mice, housed in four separate barrier rooms within the facility, fed the same chow (a VRF1 diet, SDS). The five mice sampled in each room were housed in separate cages. The five mice from each of the four rooms were taken from separate cages i.e. no two mice came from the same cage. Mice in this study were handled by individuals wearing gloves for cage cleaning purposes on a weekly basis. Mice were not housed exclusively with litter mates, with 27 individuals housed per room. Samples consisted of individual faecal pellets taken from individual mice. After collection, pellets were placed into separate collection tubes and frozen prior to analysis.

Microbiota. Nucleic acid extractions were carried out using a combination of physical disruption and phenol/chloroform extraction methods, described previously[10]. 16S rRNA gene universal Bacterial primers 27F-519R (27F 5′-AGRGTTTGATCMTGGCTCAG, 519R 5′-GTNTTACNGCGGCKGCTG) were used in a single-step 30 cycle PCR using HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA) performed under the following conditions: 94oC for 5 minutes, followed by 28 cycles of: 94oC for 30 seconds, 53oC for 40 seconds, and 72oC for 1 minute. Amplification was followed by a final elongation step at 72oC for 5 minutes. Following PCR, all amplicon products from different samples were mixed in equal concentrations and purified using Agencourt Ampure beads (Agencourt Bioscience Corporation, MA, USA). Samples were sequenced utilizing Roche 454 FLX titanium instruments and reagents following manufacturer's guidelines. A total of 165,934 16S rRNA gene sequences were obtained from the 20 faecal sample extracts. Following curation, an average of 4,356 sequences was obtained for each of the samples. For analysis of alpha and beta diversity, samples were normalised to 2,179 sequences per sample.

Sequence data analysis was carried out. Here, the Q25 sequence data derived from the sequencing process was processed using standard analysis pipeline processes (MR DNA, Shallowater, USA). Sequences were depleted of barcodes and primers then short sequences <200 bp removed, as were sequences with ambiguous base calls removed, and sequences with homopolymer runs exceeding 6 bp, sequences were denoised and chimeras removed[11–17]. Operational taxonomic units were defined after removal of singleton sequences, clustering at 3% divergence (97% similarity). Final OTUs were taxonomically classified using BLASTn against a curated databased derived from GreenGenes, NCBI and RDP databases[18]. Normalized and de-noised files were then rarefied and run through QIIME[19] to generate alpha and beta diversity data. Additional statistical analyses were performed with NCSS2007 (NCSS, UT) and XLstat 2012 (Addinsoft, NY).

A range of diversity and richness measures were used to assess changes in microbiota composition, including taxa richness, Chao1, Shannon index, Simpson index 1-D[20]. Analysis of microbiota diversity was performed using PAST - Palaeontological Statistics, version 3.01, a program available from the University of Oslo website link (http://folk.uio.no/ohammer/past).

^{1}H NMR metabolomics. Portions of mouse faeces of approximately 0.02 g were resuspended by vortexing in 500 µl of phosphate buffered saline. Particulate matter was pelleted by centrifugation at 13,000 × g for 10 min, and supernatant transferred to a fresh microfuge tube. Centrifugation was repeated, with pelleted material again discarded. Supernatant was frozen by immersion in liquid nitrogen, lyophilised at −58°C overnight, and re-suspended in 500 µl D_2O. ^{1}H NMR spectra of three replicates were acquired at 400 MHz on a Bruker Avance spectrometer (Bruker, Coventry, UK) equipped with a 5 mm QNP probe using a zgesgp pulse sequence incorporating water suppression via excitation sculpting with gradients. The ^{1}H 90 degree pulse was 9.75 µs. The spectral width was 20 ppm. Free induction decays were multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. The spectra were Fourier transformed and calibrated to a 2,2,3,3,-D4-3-(Trimethylsilyl) propionic acid (TSP) reference signal at 0 ppm. Phase correction was performed manually and automatic baseline correction was applied. To help in the assignment of the metabolite resonances, J-resolved 2D correlation was performed with pre-saturation during relaxation delay using gradients (J-Res, Bruker). Pre-processing and orthogonal projection to latent structures discriminant analysis (OPLS-DA) were carried out with software that was developed in our laboratory for a previous study[21] using the python programming language with numpy and scipy for calculations, and matplotlib for visualization. The nonlinear iterative partial least-squares (NIPALS) algorithm[22] was used for OPLS-DA analysis. Regions above 8.5 ppm and below 0.45 ppm were excluded because of noise content. The water peak and TSP reference signal were also excluded. Spectra were bucketed using 0.005 ppm bin size leaving 1588 data points per spectrum. These spectra were normalized[23,24], and auto-scaled (variance of every data point normalized to 1). Cross-validation was performed where 75% of the samples were used as a training set and the remaining 25% as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class, and that at least one sample from each class was present in the test set. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1 used to choose the number of components, with the additional constraint to use a maximum of 8 components. A double cross-validation was repeated 2000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This led to 4 × 2000 models. Finally, this procedure was repeated with randomly generated class assignments to provide a reference value for Q^2. The chosen number of components minus one was then used as an OPLS filter, and a PLS-DA analysis with two components was carried out on the filtered data to yield one predictive and one orthogonal component. In the back-scaled loadings analysis, peaks that allow the models to distinguish between classes were assigned by comparing chemical shift values and multiplicities from J-resolved NMR spectra to values from the BMRB[25] and HMDB[26].

Results & Discussion

Analysis of the bacterial identities derived from 16S ribosomal RNA gene sequencing revealed the faecal microbiota to be dominated by the phyla Bacteroidetes and Firmicutes, (62.4 ± 22.4 (SD)% and 34.7 ± 23.9%, respectively) although marked variation was observed in phylum relative abundance between individual animals (Fig. S1). Further, microbiota alpha diversity, as assessed by rarefaction and Chao1 richness estimate, OTU richness, and Shannon Index were significantly lower for mice of one room group (room 4) compared with mice from other room groups (Table S1) (Kruskall-Wallis controlled multiple pair-wise comparison, $p < 0.001$).

Analysis of microbiota at the genus level identified the twenty genera with the highest mean relative abundance (Table S2), which were broadly in keeping with those reported in the murine gut previously[27]. The most commonly numerically dominant genus was *Prevotella*, (39.0 ± 20.2% of sequences), a genus associated with a long term carbohydrate-rich diet in humans[28]. Again however, significant differences in the microbiota were identified between room groups (controlled ANOVA tests, $p < 0.05$) (Table S3). *Coprococcus*, *Ruminococcus*, and *Anaerotruncus* were significantly higher in room 1 samples, *Pedobacter* was significantly higher in room 2 samples, *Novispirillum* was significantly higher in room 3 samples and *Prevotella* was significantly higher in room 4 samples. Samples from rooms 2 and 3 groups had significantly higher abundance of *Parabacteroides* and *Sphingobacterium* than samples from rooms 1 and 4.

Hierarchal cluster analysis based upon the predominant genera indicates divergence in the composition of the microbiota into three clusters (Fig. 1). Cluster I comprised samples from all animals from room 3 and additional animals from rooms 1 and 2, cluster II comprised all animals from room 4 and cluster III included all of the remaining animals from rooms 1 and 2. Notably is the absence, or very low abundance, in room group 4 of a number of genera including *Sutterella*, *Sphingobacterium*, *Novispirillum* and *Porphyromonas*. Overall therefore, the bacterial microbiota showed marked divergence that was in cases linked to room occupancy, with these compositional differences resolving into three clusters.

Whilst all mice received the same standard diet, the differences in constituency of their microbiota indicated a potential for distinct metabolomic characteristics. The major constituent of mouse chow, carbohydrates, are fermented in the colon to short chain fatty acids (SCFA), primarily acetate, butyrate, lactate and propionate[29,30]. Whilst SCFAs are just one class of compounds, they are important in shaping the microbial community and preventing the growth of pathogens[31,32]. Moreover, SCFA levels impact on the host and are known to be important in relation to nutrition, adipose tissue deposition, immunity and cancer amongst other conditions[30,33]. Different SCFAs have been associated with effects on specific physiological processes[34], with the type of SCFAs varying between bacterial genera[35]. To test for a functionally-distinct signal, we performed a metabolomic analysis of the faecal material.

^{1}H NMR spectroscopy was performed on buffered saline extracts from the same faecal samples used for microbiota sequencing. We hypothesised that there would be differences when comparing the metabolome of faeces from mice whose faecal microbiota were distinct. Analysis involved a series of pairwise orthogonal partial least squares discriminant analysis (OPLS-DA) tests using classes suggested by clustering according to microbiota (Fig. 2), room occupancy or dominant phyla. Scores plots for each of three pairwise comparisons show that there are substantial differences in the metabolomes extracted from faeces of mice assigned to each cluster (Fig. 2 - left panels). Q^2 obtained for each test performed were compared with a reference value for Q^2, obtained after repeating cross-validation with randomly generated class assignments (Table 1). As shown, Q^2 scores for the metabolomic data pairwise analysis performed when separated according to these clusters were >0.50 which is an accepted threshold for a "good" model[36,37]. As such, we observed clear metabolomic differences in the murine faecal samples based on clusters as defined by the composition of the bacteria present. Further, microbiota data were used to assess the relative contribution of Bacteroidetes, Firmicutes and Proteobacteria to each of the samples tested. Here, Q^2 scores were all >0.41. Significant differences were also identified in the metabolome of faeces from mice housed in different room groups with Q^2 scores all >0.67 (room 2 vs. room 3).

Next, we identified the key drivers of the differences in the metabolomic data by generating back-scaled loadings plots and assigning resonances with high variance and high weight, indicated by greater intensity and yellow/red color respectively (Fig. 2 – right panels). Notably, Clusters I and II were distinguished by the greater abundance of a number of amino acids in the faecal metabolomes of mice in Cluster II whereas the faecal metabolomes of mice from Cluster III were distinguished from those in Cluster I and II on the basis of short chain fatty acids which were more abundant in Cluster III. At the outset of this study, we hypothesised that there would be minimal differences between the gut microbiota as sampled in the context of genetically identical mice. However, significant differences were observed in the taxa detected, their relative abundance, and overall bacterial diversity. This variation in the faecal microbiota was linked, at least in part, to the barrier room in which the mice were housed. Assessment of the metabolome associated with these animals showed that microbiota and metabolome findings were largely consistent.

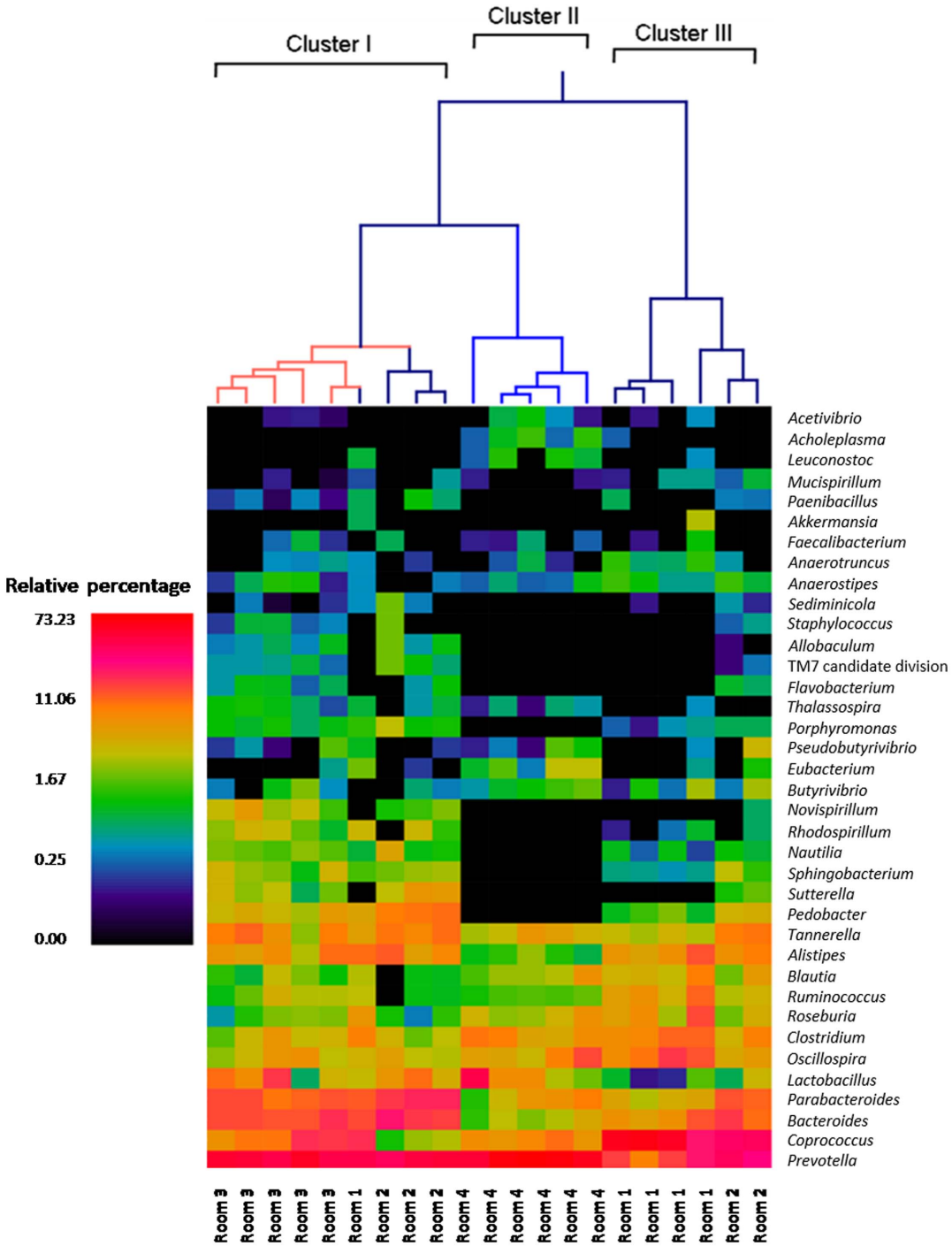


Figure 1 | Heat map analysis of the predominant genera identified in this study. A hierarchal cluster diagram was constructed using Ward's minimum variance clustering and Manhattan distances. Room group 1 and room group 2 exhibit some co-clustering indicating differences within the groups. The heatmap describes the relative percentage in each sample of the associated genera with a legend provided in the upper left of the figure.

Murine models are used in biomedical research to address almost every aspect of human health. To avoid potentially confounding differences in genetic backgrounds, mice are taken from inbred populations with the rationale being that the resulting homogeneity provides a uniform "platform" for study. By far the most common genetic background for mice used as models of human disease is the strain C57BL/6, as used here. When purchased for research, individual C57BL/6 mice are commonly considered to be equivalent. Increasingly however, the potential of the gastrointestinal microbiota to influence the host in relation to health and a wide range of clinical syndromes is being recognised[38]. In this light, the differences identified in microbiota here require further consideration. Given the potential impact of the gut microbiota on so many important physiological processes, the degree to which it is conserved between individual animals used in biological research is arguably as important as their genetic uniformity. Further, variation in gut microbiota composition is likely to be even higher in less well controlled experimental facilities, and to be exacerbated when mice are moved between facilities, experience changes in diet, and are exposed to animals with different microbiota. The divergence in gut microbiota

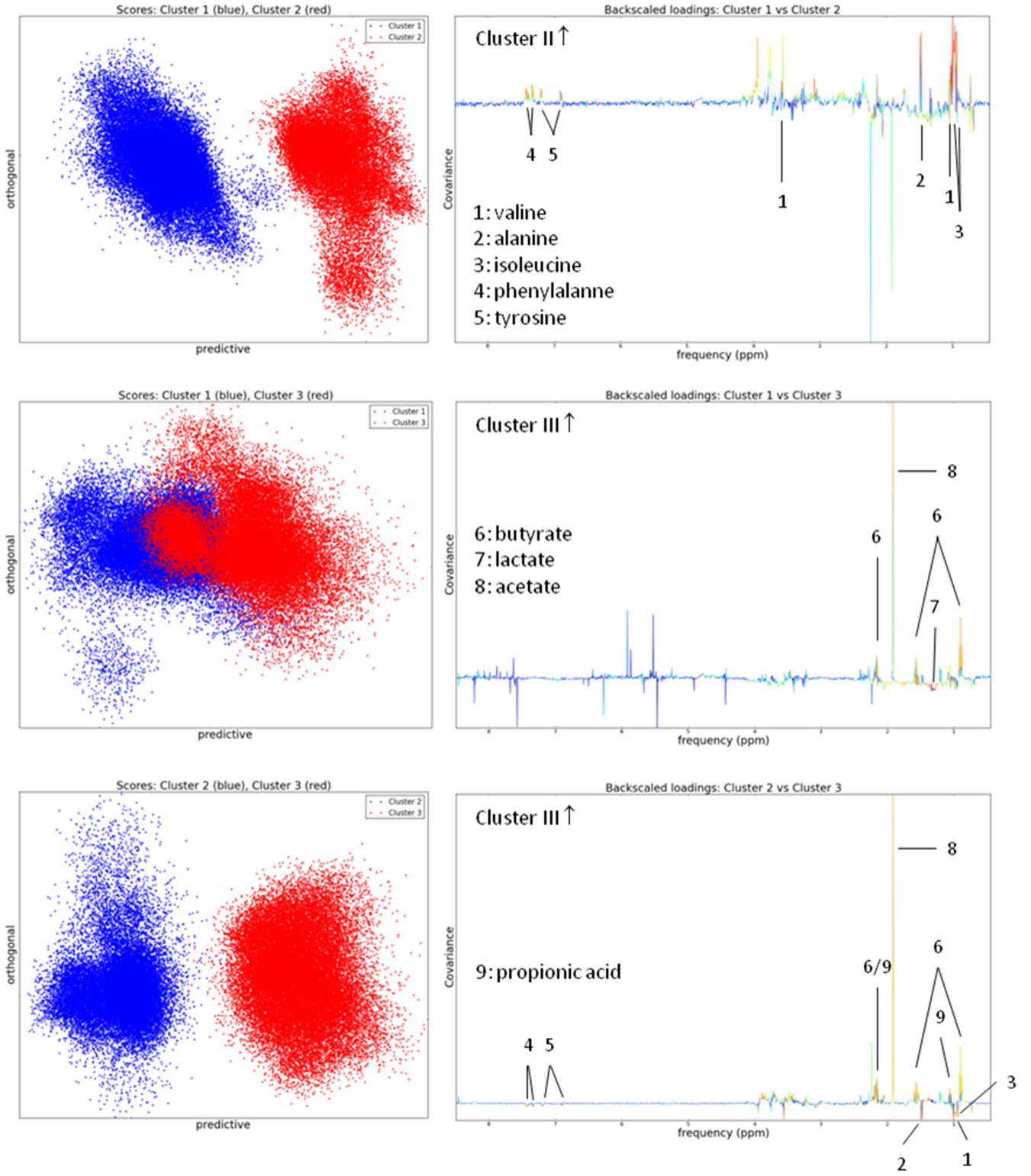


Figure 2 | OPLS-DA scores plots (left panels) and back-scaled loadings plots (right panels) for comparisons between the murine faecal metabolomes as clustered according to microbiota data composition. Resonances with high variance and high weight are highlighted in red. The distinguishing metabolites that could be unambiguously assigned are annotated in each back-scaled loadings plot. Q^2 values for the cross-validated OPLS-DA comparisons are provided in Table 1.

Table 1 | Predictive Q^2 values for all models. Q^2 values for models run with permutated class assignments are given in parentheses

Model	Q^2
Cluster I vs Cluster II	0.88 (−0.15)
Cluster I vs Cluster III	0.52 (−0.15)
Cluster II vs Cluster III	0.81 (−0.18)
Room 1 vs Room 2	0.93 (−0.14)
Room 1 vs Room 3	0.90 (−0.15)
Room 1 vs Room 4	0.85 (−0.15)
Room 2 vs Room 3	0.67 (−0.09)
Room 2 vs Room 4	0.80 (−0.12)
Room 3 vs Room 4	0.86 (−0.15)
High Bacteroidetes vs low Bacteroidetes	0.41 (−0.15)
High Firmicutes vs low Firmicutes	0.41 (−0.17)
High Proteobacteria vs low Proteobacteria	0.66 (−0.18)

composition, as reflected in faecal bacteria, strongly suggests that efforts must be made to ensure uniformity of intestinal microbiota in animals used in research.

1. Bäckhed, F. Changes in intestinal microflora in obesity: cause or consequence? *J. Pediatr. Gastroenterol. Nutr.* **48** Suppl 2, S56–7 (2009).
2. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–31 (2010).
3. Stepankova, R. *et al.* Absence of microbiota (germ-free conditions) accelerates the atherosclerosis in ApoE-deficient mice fed standard low cholesterol diet. *J. Atheroscler. Thromb.* **17**, 796–804 (2010).
4. Diaz Heijtz, R. *et al.* Normal gut microbiota modulates brain development and behavior. *Proc. Natl. Acad. Sci. U S A* **108**, 3047–52 (2011).
5. Cani, P. D. *et al.* Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* **57**, 1470–81 (2008).
6. Claus, S. P. *et al.* Colonization-induced host-gut microbial metabolic interaction. *MBio* **2**, e00271-10; DOI:10.1128/mBio.00271-10 (2011).
7. Björkholm, B. *et al.* Intestinal microbiota regulate xenobiotic metabolism in the liver. *PLoS One* **4**, e6958; DOI:10.1371/journal.pone.0006958 (2009).
8. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U S A.* **104**, 13780–5 (2007).
9. Lee, S. H. *et al.* ERK activation drives intestinal tumorigenesis in Apc(min/+) mice. *Nat. Med.* **16**, 665–70 (2010).
10. Rogers, G. B. *et al.* A Novel Microbiota Stratification System Predicts Future xacerbations in Bronchiectasis. *Ann. Am. Thorac. Soc.* Epub ahead of print; DOI:10.1513/AnnalsATS.201310-335OC (2014).
11. Dowd, S. E. *et al.* Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol.* **8**, 125 (2008).
12. Dowd, S. E., Sun, Y., Wolcott, R. D., Domingo, A. & Carroll, J. A. Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: bacterial diversity in the ileum of newly weaned *Salmonella*-infected pigs. *Foodborne Pathog. Dis.* **5**, 459–72 (2008).
13. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–1 (2010).
14. Capone, K. A., Dowd, S. E., Stamatas, G. N. & Nikolovski, J. Diversity of the human skin microbiome early in life. *J. Invest. Dermatol.* **131**, 2026–32 (2011).
15. Dowd, S. E. *et al.* Survey of fungi and yeast in polymicrobial infections in chronic wounds. *J. Wound Care* **20**, 40–7 (2011).
16. Eren, A. M. *et al.* Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS One* **6**, e26732; DOI:10.1371/journal.pone.0026732 (2011).
17. Swanson, K. S. *et al.* Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J.* **5**, 639–49 (2011).
18. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–72 (2006).
19. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–6 (2010).
20. Haegeman, B. *et al.* Robust estimation of microbial diversity in theory and in practice. *ISME J.* **7**, 1092–101 (2013).
21. Vermeer, L. S., Fruhwirth, G. O., Pandya, P., Ng, T. & Mason, A. J. NMR metabolomics of MTLn3E breast cancer cells identifies a role for CXCR4 in lipid and choline regulation. *J. Proteome Res.* **11**, 2996–3003 (2012).
22. Andersson, M. A comparison of nine PLS1 algorithms. *J. Chemometrics* **23**, 518–529 (2009).
23. Kozlowska, J. *et al.* The impact of Pseudomonal growth on cystic fibrosis airway secretion composition in a metabolomic investigation. *Metabolomics.* **9**, 1262–1273 (2013).
24. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **78**, 4281–90 (2006).
25. Ulrich, E. L. *et al.* BioMagResBank. Nucleic Acids Res. **36**, D402–8 (2008). Wishart, D. S. Computational approaches to metabolomics. *Methods Mol. Biol.* **593**, 283–313 (2010).
26. Wishart, D. S. Computational approaches to metabolomics. *Methods Mol. Biol.* **593**, 283–313. (2010).
27. Linnenbrink, M. *et al.* The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol. Ecol.* **22**, 1904–16 (2013).
28. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–8 (2011).
29. Roberfroid, M. *et al.* Prebiotic effects: metabolic and health benefits. *Br. J. Nutr.* **104** Suppl 2, S1–63 (2010).
30. Delzenne, N. M. & Cani, P. D. Interaction between obesity and the gut microbiota: relevance in nutrition. *Annu. Rev. Nutr.* **31**, 15–31 (2011).
31. Blaut, M. Relationship of prebiotics and food to intestinal microflora. *Eur. J. Nutr.* **41** Suppl 1, I11–6 (2002).
32. Duncan, S. H., Louis, P., Thomson, J. M. & Flint, H. J. The role of pH in determining the species composition of the human colonic microbiota. *Environ. Microbiol.* **11**, 2112–22 (2009).
33. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–6 (2009).
34. Xiong, Y. *et al.* Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. *Proc. Natl. Acad. Sci. U S A.* **101**, 1045–50. (2004).
35. Hague, A. *et al.* Sodium butyrate induces apoptosis in human colonic tumour cell lines in a p53-independent pathway: implications for the possible role of dietary fibre in the prevention of large-bowel cancer. *Int. J. Cancer* **55**, 498–505 (1993).
36. Szymanska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, S3–S16 (2012).
37. Westerhuis, J. A. *et al.* Assessment of PLSDA cross validation. *Metabolomics* **4**, 81–9 (2008).
38. Shanahan, F. The colonic microbiota in health and disease. *Curr. Opin. Gastroenterol.* **29**, 49–54 (2013).

Author contributions

G.B.R., K.D.B., A.J.M. and M.M. wrote the main manuscript text and prepared all figures. S.E.D. and G.B.R. prepared figure 1 A.J.M. prepared figure 2. J.Ko. performed 1H NMR, J.Ke. K.M. and M.F. were responsibility for animal housing and sample collection, and S.E.D. performed 16S rRNA gene sequencing. All authors approved the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Rogers, G.B. *et al.* Functional divergence in gastrointestinal microbiota in physically-separated genetically identical mice. *Sci. Rep.* **4**, 5437; DOI:10.1038/srep05437 (2014).

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Combined Systems Approaches Reveal Highly Plastic Responses to Antimicrobial Peptide Challenge in *Escherichia coli*

Justyna Kozlowska[1], Louic S. Vermeer[1], Geraint B. Rogers[1¤], Nabila Rehnnuma[1], Sarah-Beth T. A. Amos[1], Garrit Koller[2], Michael McArthur[3], Kenneth D. Bruce[1], A. James Mason[1]*

1 King's College London, Institute of Pharmaceutical Science, London, United Kingdom, **2** King's College London, King's College London Dental Institute at Guy's, King's and St. Thomas' Hospitals, London, United Kingdom, **3** Department of Molecular Microbiology, John Innes Centre, Norwich, United Kingdom

Abstract

Obtaining an in-depth understanding of the arms races between peptides comprising the innate immune response and bacterial pathogens is of fundamental interest and will inform the development of new antibacterial therapeutics. We investigated whether a whole organism view of antimicrobial peptide (AMP) challenge on *Escherichia coli* would provide a suitably sophisticated bacterial perspective on AMP mechanism of action. Selecting structurally and physically related AMPs but with expected differences in bactericidal strategy, we monitored changes in bacterial metabolomes, morphological features and gene expression following AMP challenge at sub-lethal concentrations. For each technique, the vast majority of changes were specific to each AMP, with such a plastic response indicating *E. coli* is highly capable of discriminating between specific antibiotic challenges. Analysis of the ontological profiles generated from the transcriptomic analyses suggests this approach can accurately predict the antibacterial mode of action, providing a fresh, novel perspective for previous functional and biophysical studies.

Citation: Kozlowska J, Vermeer LS, Rogers GB, Rehnnuma N, Amos S-BTA, et al. (2014) Combined Systems Approaches Reveal Highly Plastic Responses to Antimicrobial Peptide Challenge in *Escherichia coli*. PLoS Pathog 10(5): e1004104. doi:10.1371/journal.ppat.1004104

Editor: Jeff H. Chang, Oregon State University, United States of America

Received January 9, 2014; **Accepted** March 18, 2014; **Published** May 1, 2014

Funding: This work was supported by the Medical Research Council (http://www.mrc.ac.uk/index.htm, NIRG G0801072/87482 to AJM), the Biotechnology and Biological Sciences Research Council (http://www.bbsrc.ac.uk/home/home.aspx) and Procarta Biosystems Ltd (http://www.procartabio.com/, Industrial CASE award BB/H01571X/1 supporting JK) and the Wellcome Trust (http://www.wellcome.ac.uk/, Capital Award for the KCL Centre for Biomolecular Spectroscopy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: james.mason@kcl.ac.uk

¤ Current address: SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia

Introduction

The isolation of cecropins [1], magainins [2] and defensins [3] from insects, amphibians and mammals in the late 1980's and early 1990's, highlighted the potential of host defence peptides as sources of novel antibiotics [4]. This novel antibiotic potential encouraged researchers to develop structure activity relationships for cationic antimicrobial peptides (AMPs), with the anionic bacterial plasma membrane the presumed site of action for bactericidal activity [5]. There is increasing evidence however that each AMP may indeed have multiple effects on a bacterial cell and hence may have multiple ways of killing microbial targets. AMPs may therefore function as "dirty drugs" with different bactericidal strategies possible for distinct bacterial species [4–7]. Indeed, the innate immune system may have selected AMPs that can exert their antimicrobial activity in multiple ways since this is less likely to lead to resistance developing as seen with classical antibiotics that have a single, high affinity target [6]. Our understanding of how AMPs function is therefore far from complete. Attempts to optimize AMP potency in the laboratory, that focus on only one possible bactericidal mechanism, ignore the possibilities offered by taking a holistic approach that can reveal the true source(s) of bactericidal potency along with a better understanding of bacterial counter-measures.

The full power of 'omics based research tools has yet to be brought to bear in antibiotic research [8]. Nevertheless, important insights have emerged regarding the scope of bacterial responses by comparing challenges with distinct AMPs [8]. These studies have focussed on the Gram-positive bacterial species *Bacillus subtilis* [9], *Staphylococcus aureus* [10] and *Streptococcus pneumoniae* [11] and have demonstrated the existence of complex regulatory patterns in which several signal transduction pathways were induced. The transcriptional response of *Escherichia coli* to cecropin A, the proline rich Bac7(1-35) and novispirin G10 has been characterised in separate studies [12–14]. Recent work in our laboratory has focussed on trying to understand the relative difference in antibacterial potency of structurally related AMPs to Gram-negative bacteria such as *Escherichia coli* and *Pseudomonas aeruginosa* [15–18]. Here, AMPs with structural features thought to enhance antibacterial potency and reduce toxicity have been developed for use against more challenging pathogens [19,20]. These peptides, including D-LAK120-AP13, have been developed based on an understanding of a variety of naturally occurring peptides including magainin 2, buforin II and pleurocidin. Pleurocidin is a 25 amino acid AMP found in the skin and gills of *Pleuronectes americanus*, the Winter Flounder. Despite resembling magainin 2 in terms of length, cationic charge, hydrophobicity and secondary

Author Summary

Antimicrobial peptides (AMP) are small proteins with often potent antibacterial activity found in a variety of organisms, including humans. Understanding how these antibiotics operate is challenging and often controversial since many studies have necessarily focussed on identifying a single major cause of bacterial cell death while, increasingly, others have cautioned that AMPs are likely to have access to multiple bactericidal features. Systems biology is an emerging field that comprises a series of techniques capable of giving a global view of how bacteria respond to external stimuli. Here we have monitored changes in gene expression and metabolism in bacteria that have been challenged with sub-lethal concentrations of four different AMPs. By understanding how bacteria respond to a threat we can reveal how the bacteria perceive the AMP to be operating. Our approach provides a sophisticated bacterial perspective of the mode of action of each AMP and reveals that the bacteria have a vast array of weapons that can be marshalled to deal with distinct AMP threats. Indeed, around a third (or even more) of the bacterial machinery might be useful in dealing with antibiotic challenges, highlighting why antibiotic resistance is such a persistent problem.

structure in a range of membrane mimetic environments [18], pleurocidin is typically ten times more potent against Gram-negative species. Pleurocidin has been shown to be capable of acting on bacterial membranes [21], with pore forming activity, but has also been suggested to enter bacterial cells and interrupt protein synthesis [22]. We have therefore compared its effect on *E. coli* with magainin 2, which has been considered the archetypal pore forming AMP, and with buforin II which is proposed to enter bacteria to exert a bactericidal effect [23,24].

Since these peptides act at widely differing effective concentrations we hypothesised that studying their effects at sub-lethal concentrations would provide a detailed overview of the mechanisms of action of each AMP. We therefore devised a method that could efficiently identify conditions where bacteria responded to AMP challenge without introducing possible, non-specific complications that might result from large scale cell death. We therefore used ^{1}H high resolution magic angle spinning (HR-MAS) NMR to identify the lowest AMP concentration that elicited a response from metabolically active, challenged bacteria. A robust, cross-validated, multivariate analysis identified metabolites whose levels were altered in response to AMP challenge. These were used to classify the AMP according to the elicited response whilst providing a first indication of whether *E. coli* responded in a generic or specific manner to AMP challenge. Having identified sub-lethal conditions where a response was confirmed, electron microscopy and transcript profile analyses enabled a detailed description of the *E. coli* response to AMP challenge.

Results

^{1}H HR-MAS NMR metabolomics reveals threshold AMP concentration inducing a bacterial response –The four cationic amphipathic AMPS selected to test the response of stationary phase *E. coli* (Table 1) were of similar length and were all C-terminally amidated with nominal charge ranging from +4 to +9. For the AMP challenge experiments presented here, higher bacterial cell densities (8×10^{8} CFU/ml) were required than is common in the broth microdilution assays [25] used to generate

Table 1. Sequences of peptides used in this study.

Peptide	Sequence	Length	Charge	Average Hydro-phobicity (*H*)*	Hydrophobic moment (μ*H*)	MIC (μg/ml) *E. coli* NCTC 9001	MIC (μg/ml) *P. aeruginosa* PAO1
Magainin 2	GIGKFLHSAKKFGKAFVGEIMNS	23	+4	−0.03	0.28	16.91±3.89	26.12±5.45
Pleurocidin	GWGSFFKKAAHVGKHVGKAALTHYL	25	+5	−0.02	0.22	1.79±0.60	4.47±2.36
Buforin II	TRSSRAGLQFPVGRVHRLLRK	21	+7	−0.37	0.30	>64.00	>64.00
D-LAK120-AP13	*KKLALALAKKWLPLAKKLALALAKK*	25	+9	−0.10	0.21	2.95±1.69	3.33±0.52

MIC data are as reported in (17) and (18).
*As determined by the Eisenberg (1982) Consensus scale (Ile, 0.73; Phe, 0.61; Val, 0.54; Leu, 0.53; Trp, 0.37; Met, 0.26; Ala, 0.25; Gly, 0.16; Cys, 0.04; Tyr, 0.02; Pro, −0.07; Thr, −0.18; Ser, −0.26; His, −0.40; Glu, −0.62; Asn, −0.64; Gln, −0.69; Asp, −0.72; Lys, −1.1; and Arg, −1.8).
doi:10.1371/journal.ppat.1004104.t001

the MIC data (Table 1), in particular for HR-MAS analysis. At the higher bacterial titre, although the relative potency is similar, the effect of the four AMPs determined using such methods was somewhat different from the minimum inhibitory concentrations (MICs), with D-LAK120-AP13 having a substantially greater effect on bacterial numbers as detected in the challenge and recovery assay (Fig. 1A). Neither magainin 2 nor buforin II had a sufficiently inhibitory effect for a MIC to be determined at the higher titre. Nevertheless the amount of peptide causing a significant reduction in bacterial re-growth can be compared with D-LAK120AP13 effective at 15.6 μg/ml with substantially more pleurocidin (62.5 μg/ml) and magainin 2 (125 μg/ml) required for a significant effect. No effect on bacterial re-growth was observed for buforin II at any of the peptide concentrations tested (Fig. 1A). A multi-parameter assay was taken to assess the effect of peptide challenge on membrane potential (Fig. 1B), esterase activity (Fig. 1C) and membrane integrity in the challenged stationary phase bacteria and suggested that only the higher concentrations of pleurocidin and D-LAK120-AP13 were lethal. A dose dependent response to each of the four AMPs was observed but the membrane potential was not completely lost while the esterase activity was mostly higher than that observed in untreated cells; a hallmark of exposure to sublethal stress in *E. coli* [26].

One dimensional ^{1}H NMR spectra were obtained for all samples. Principal component analysis identified outlier spectra resulting from either poor baseline or signal to noise, and either partial least squares (PLS) regression analysis or a series of orthogonal PLS-discriminant analysis (OPLS-DA) tests were used to interrogate the spectra. The latter was used in a step-wise manner to determine the lowest concentration of each AMP that caused a significant change in the spectra relative to spectra from untreated bacterial cell, as determined from Q^2 (Table 2) where an arbitrary value $\geq$0.6 was taken to show a reliable model where the AMP challenge has a significant effect. This value can be compared with the value expected for a perfect separation of the two groups ($Q^2 = 1$) and that obtained when the assigned classes (untreated or AMP challenged) are permutated as a means of representing no effect. 2D scores plots that resulted from each of the cross validated OPLS-DA analyses are shown in the supplementary material (Fig. S1 in File S1) while those for the threshold concentrations (data for the highest AMP concentration is shown for buforin II) are described here (Fig. 2A–D). A clear separation of the OPLS-DA scores was obtained at the indicated AMP concentration, identified as a threshold for a response detectable in the bacterial metabolomes. This threshold concentration varied considerably for the four AMPs and was directly related to the apparent antibacterial efficacy noted above. The threshold value, in turn, corresponded to a sub-lethal AMP concentration since bacterial growth remained greater than 50% of maximum (Fig. 1A), esterase activity was increased relative to untreated cells (Fig. 1C) while membrane potential was not completely lost (Fig. 1B). Hence the NMR metabolomic technique identified conditions where metabolically active *E. coli* were responding to the AMP challenge without simply reporting on bacterial cell death. Comparing the back-scaled loadings, each of the OPLS-DA comparisons between untreated bacteria and those challenged with each AMP, identified metabolites whose differing intensities correlated with the effect of each AMP. A hierarchical cluster analysis was used to reveal variation in metabolite levels (Fig. 2E). Both common and AMP specific variations in *E. coli* metabolite levels were observed in response to challenge with the four AMPs. Notably, the hierarchical analysis grouped the peptides according to their potency. Though not considering the magnitude or direction of changes in metabolite levels, network

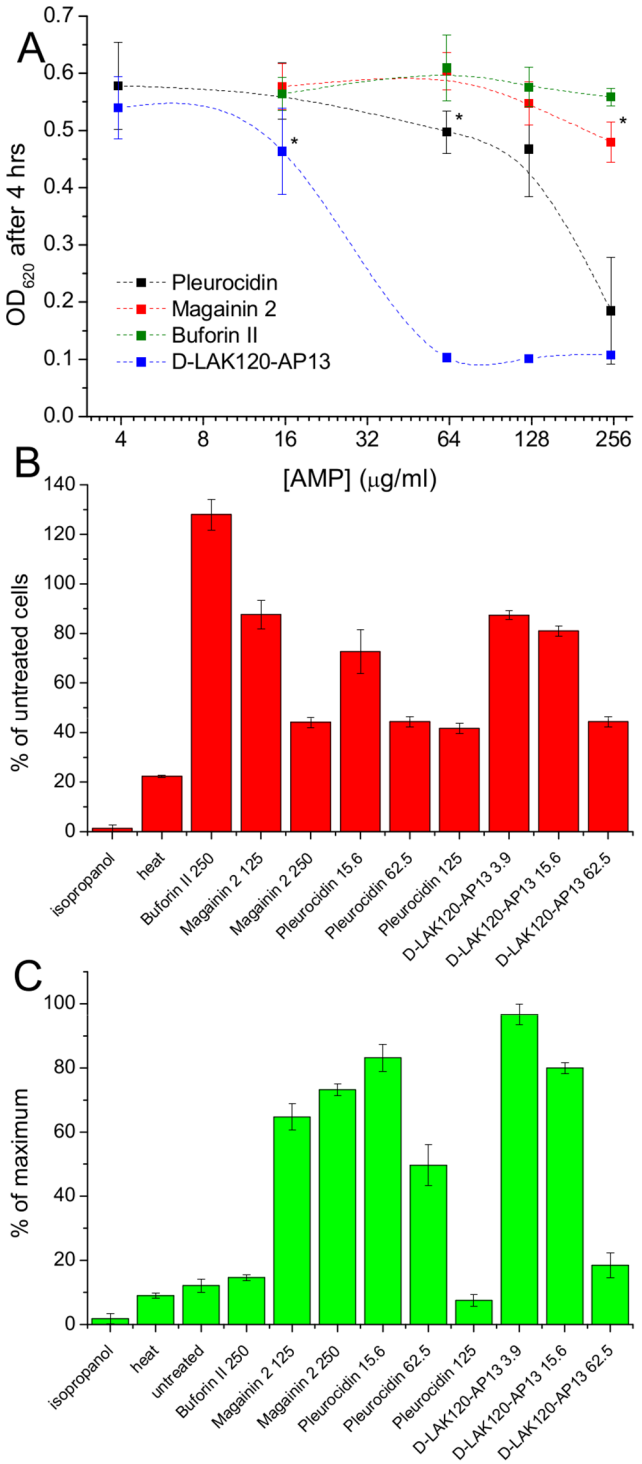


Figure 1. AMP challenge and multi-parameter assay of *E. coli* NCTC 9001. Overnight cultures were challenged with increasing amounts of each of four AMPs for 30 minutes and the recovery of aliquots added to fresh media was measured after 4 hours incubation at 37°C (**A**). * indicates the peptide concentration causing a significant ($p<0.1$) reduction in OD_{620} relative to the lowest peptide concentration used. The membrane potential (**B**) of challenged bacteria as measured by the voltage sensitive dye $DiBAC_4$ is expressed here as a percentage of the membrane potential determined for untreated cells. Esterase activity (**C**) determined by cleavage of 5,6-carboxyfluorescein diacetate expressed as a percentage of the maximum observed activity. Peptide concentrations are given in μg/ml.
doi:10.1371/journal.ppat.1004104.g001

Table 2. Predictive Q^2 values for OPLS-DA models.

[AMP] (μg/ml)	Q^2			
	Pleurocidin	**Magainin 2**	**Buforin II**	**D-LAK120-AP13**
3.9	0.32 (−0.29)	n.d	n.d.	0.37 (−0.30)
15.6	0.53 (−0.31)	0.29 (−0.36)	n.d.	0.59 (−0.28)*
62.5	0.81 (−0.41)*	0.20 (−0.31)	n.d.	0.81 (−0.31)
125	0.80 (−0.29)	0.68 (−0.34)*	n.d.	0.83 (−0.26)
250	n.d.	n.d.	−0.30 (−0.39)*	n.d.

Q^2 values for cross validation performed with permutated classes are provided in parentheses.
* Key minimum concentrations.
doi:10.1371/journal.ppat.1004104.t002

pathway analysis conducted using MetaboAnalyst [27,28] matched pathways according to p-values obtained from pathway enrichment analysis and pathway impact from pathway topology (Fig. S2.1–S2.4 in File S1). Changes in alanine, aspartate and glutamate metabolism had the greatest impact and were a common feature of challenge with all four peptides with changes in pyruvate, butanoate and arginine/proline metabolism highlighted according to the distinct challenges.

The dynamic response of *E. coli* NCTC9001 to challenge with pleurocidin or magainin 2 was assessed over a period of 2 hours at the following intervals: 5 minutes, 15 minutes, 60 minutes and 120 minutes. The OPLS-DA scores plots (Fig. S3 in File S1) and corresponding Q^2 (Table S1 in File S1) indicate that a response to AMP challenge at the level of the metabolome can be detected throughout the period tested. However, when the back-scaled loadings were compared in a hierarchical cluster analysis (Fig. S4 in File S1), modest but notable differences in the affected metabolites were discerned. This suggested that the bacterial response detected beyond an hour after challenge is characteristically distinct from that probed within the first 30 minutes. These conditions – 30 minutes incubation at the determined threshold concentration - were therefore used for subsequent electron microscopic and transcript profiling analyses of samples prepared in parallel to those used above.

Scanning and transmission electron microscopy identifies differences in the response to each AMP

Changes in *E. coli* internal or external morphology in response to challenge with AMP were monitored respectively using transmission and scanning electron microscopy (TEM/SEM) at either one or four times the sub-inhibitory AMP threshold concentration known to induce a metabolomic response (Fig. 3; Fig. S5.1–5.10 in File S1). The bacterial response to each AMP challenge varied considerably and was in qualitative agreement with the metabolomic study; buforin II had no noticeable effect when compared with untreated bacterial cell controls (Fig. 3D–F; Fig. S5.2/5.9/5.10 in File S1), with each of the three other AMPS inducing substantial changes to external and/or internal morphologies. For magainin 2, a regular, almost circular nucleoid condensation was observed in some, but not all, cells (Fig. 3A; Fig. S5.8 in File S1) while some impairment of cell division was evident with extended rods observed (Fig. 3G). Pleurocidin also induced nucleoid condensation but this was much more widespread; observed throughout the bacterial cell population (Fig. S4.5/S4.6 in File S1). This was accompanied by some possible protein aggregation and the production of large amounts of a fibrous material (Fig. 3B). In addition to the production of the fibrous material, SEM indentified moderate vesicle production, a known envelope stress response in Gram-negative bacteria [29]. Finally, D-LAK120-AP13 induced dramatic changes in both the internal (Fig. 3C) and external *E. coli* morphologies (Fig. 3I). Extensive release of outer membrane vesicles was evident which was coincident with a loss of the normal rod shape, consistent with bacteria budding prematurely (Fig. 3I). Inside bacterial cells, extensive nucleoid condensation and protein aggregation was observed throughout the bacterial cell population (Fig. 3C; Fig. S5.3/S5.4 in File S1). Taken together, although there were some qualitative similarities in the response of *E. coli* cells to each of the three more potent AMPs, markedly distinct responses to each peptide were observed overall. Transmission electron micrographs obtained at higher magnification and with AMP added at a concentration above the detected threshold value indicated that, for all four peptides, the bacterial envelope remained intact and no release of cell contents was apparent (Fig. S5.1 in File S1).

Global transcriptome response identifies some generic, but largely AMP specific, responses

The response of *E. coli* to challenge with the four AMPs was then probed at the level of the transcriptome. Transcript profile changes in the NCTC 9001 strain, a clinical isolate from a patient with cystitis with cystitis, were monitored using the *E. coli* Genome 2.0 Array where four strains including laboratory, uropathogenic and enteropathogenic strains are featured. Due to the high degree of similarity between strains, in the majority of cases, a single probe set represents the equivalent ortholog in all four strains. All genes that are subsequently described in detail are found in both laboratory (K12 substr. MG1655) and uropathogenic (CFT073) strains with the majority also found in the two enteropathogenic strains. Principal component analysis of the twenty most differentially expressed genes across all groups showed the three independent replicates of each condition clustered together indicating the AMP challenge and transcript profiling assay were reproducible (Fig. S6 in File S1). Further analysis, where either an arbitrary significance level ($p \leq 0.05$) for differential gene expression or manual manipulation of significance levels leading to an optimal separation by principal components, generated lists of differentially expressed genes related to each treatment. *E. coli* genomes commonly encode between approximately 4,200 and 5,500 protein coding genes [30,31]. Of the approximately 10,000 probe positions, between 139 and 632 differentially expressed unique genes ($p \leq 0.05$) were detected for each treatment following challenge with AMP at the threshold concentration eliciting a bacterial response. This corresponds to 2.5–15.0% of the available genome. Magainin 2 induced differential expression of only 139

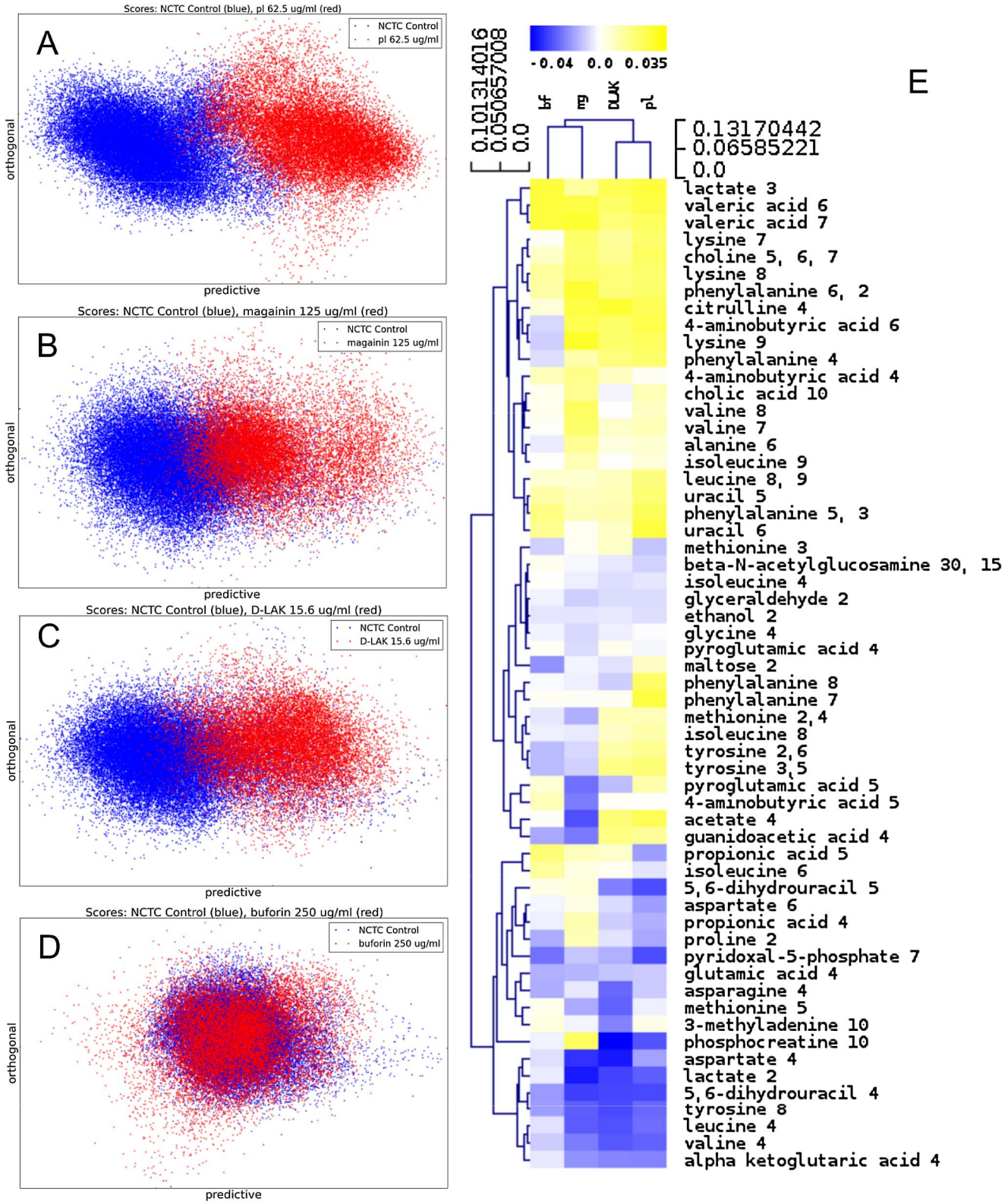


Figure 2. Metabolomic analysis by ^{1}H HR-MAS NMR of lyophilised, stationary phase *E. coli* cell pellets. OPLS-DA scores plots are shown for challenge of *E. coli* NCTC 9001 at the following threshold concentrations; pleurocidin at 62.5 μg/ml (**A**), magainin 2 at 125 μg/ml (**B**), D-LAK120-AP13 at 15.6 μg/ml (**C**) and buforin II at 250 μg/ml (**D**). Hierachical clustered heatmap comparing loadings obtained from cross-validated OPLS-DA comparing untreated bacteria with AMP at the threshold concentrations indicated above (**E**).
doi:10.1371/journal.ppat.1004104.g002

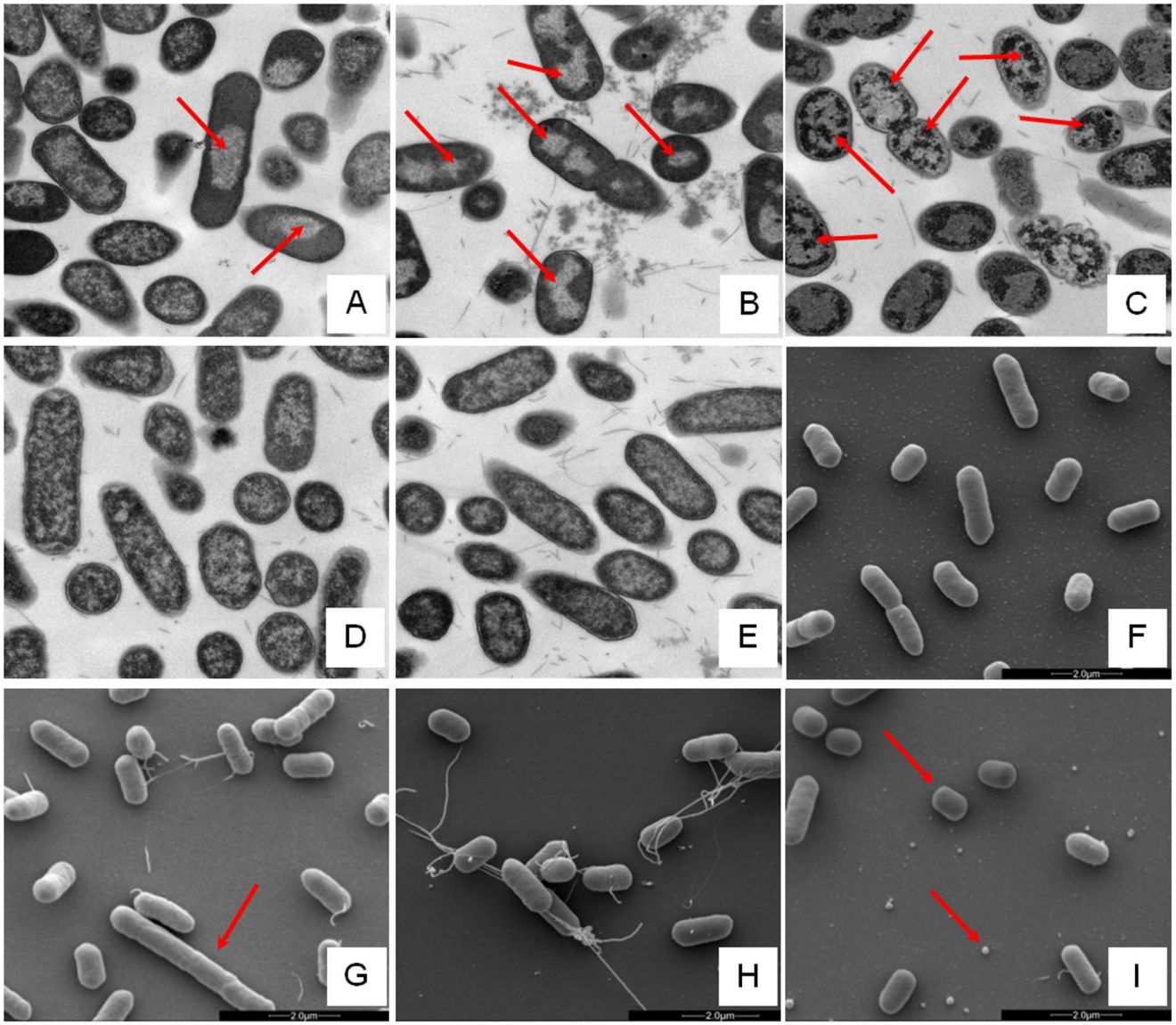


Figure 3. Electron microscopic analysis of *E. coli* response to AMP challenge. Transmission (**A–E**) and scanning (**F–I**) electron micrographs at ×25,000 magnification of either untreated (**E/F**) or AMP challenged *E. coli* NCTC 9001. Stationary phase bacteria were challenged for 30 minutes with AMPs above the threshold concentration that elicits a bacterial response as determined by the ^{1}H NMR metabolomic study; 250 μg/ml magainin 2 (**A/G**), 125 μg/ml pleurocidin (**B/H**), 62.5 μg/ml D-LAK120-AP13 (**C/I**) and 250 μg/ml buforin II (**D**). Red arrows indicate features described in the results.
doi:10.1371/journal.ppat.1004104.g003

genes which contrasted with the much greater number of genes whose expression was altered in response to challenge with either buforin II or D-LAK120-AP13; 625 and 632 respectively. Pleurocidin induced differential expression of 298 genes. The distribution of differentially expressed genes according to each AMP treatment is represented in a Venn diagram and reveals that the vast majority (76.3%) are specific to each of the four AMP challenges (Fig. 4A). Only 32 differentially expressed genes, 2.4% of the total, were common to at least three treatments while there was only one, *yjjB*, which was common to all four treatments. Qualitatively therefore, transcriptomic data supported the electron microscopy findings as, while common responses can be identified, the dominant impression was of a largely specific response to each AMP challenge

Mapping those discriminating metabolite changes with most impact (Fig. S2.1–S2.4 in File S1) onto their respective Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways identified differentially expressed genes with a potentially key role in mediating the response to AMP challenge. Changes in alanine, aspartate and glutamate metabolism were common to all four peptides and changes in expression of *gltX*, *dapA* and *metB*, coding for respectively glutamyl-tRNA synthetase, dihydropicolinate synthase and cystathionine gamma-synthase, were observed in the gene lists though these did not always satisfy the significance thresholds used above. Knockout mutants of *dapA* and *gltX* are not available from the Keio collection but Δ*metB* and five other knockout mutants (Δ*cyoA*, Δ*cyoC*, Δ*cyoD*, Δ*speB*, and Δ*argR* coding respectively for cytochrome o uniquinol oxidase subunits II, III and IV, agmatinase and arginine repressor), linked to changes in arginine/proline metabolism, were tested for altered sensitivity to AMP challenge though none was found.

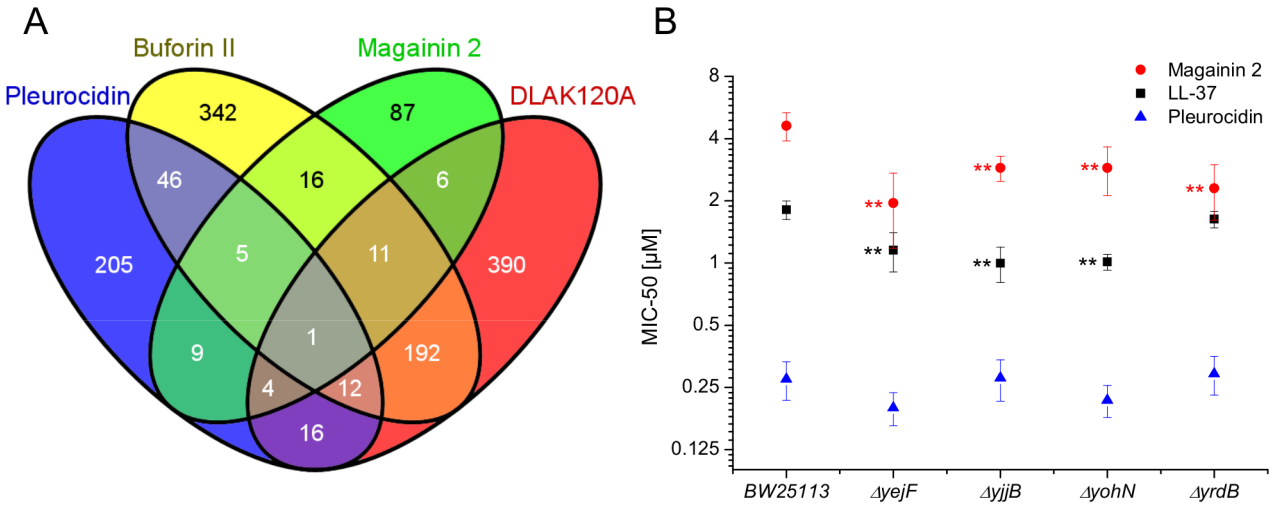


Figure 4. Transcript profiles and role of individual genes in response to AMP challenge. Four way Venn diagram (**A**) showing the distribution of differentially expressed genes detected by the GeneChip *E. coli* Genome 2.0 Array ($p \leq 0.05$) following challenge of stationary phase *E. coli* NCTC 9001 with each of four AMPs at sub-inhibitory concentrations known to elicit a bacterial response; pleurocidin at 62.5 μg/ml, magainin 2 at 125 μg/ml, D-LAK120-AP13 at 15.6 μg/ml and buforin II at 250 μg/ml. The entries in the Venn correspond to the number of affected genes. Effect on sensitivity of *E. coli* BW25113 to magainin 2, pleurocidin and LL-37 of mutations in four of six genes commonly regulated in response to AMPs of natural origin (**B**).
doi:10.1371/journal.ppat.1004104.g004

Up-regulated in response to challenge by all four AMPs, *yjjB*, encodes a 157 amino acid, conserved, inner membrane protein predicted to have four trans-membrane helices but with no known function. Of the five genes whose expression was generically affected by the three AMPs of natural origin, three were up-regulated in response to AMP challenge; *manA* codes for mannose-6-phosphate isomerise, *cysE* codes for a serine acetyltransferase and *yohN* codes for a 112 amino acid integral membrane protein annotated and established as a periplasmic modulator of nickel and cobalt efflux and renamed *rcnB* [32]. In contrast, *yejF*, part of an ABC transporter identified as a possible nickel, and probable microcin C transporter [33], and *yrdB*, which codes for a highly anionic, glutamine rich, 85 amino acid hypothetical protein from the DUF1488 superfamily, are down-regulated. Comparison of the growth of parent strain BW25113 and four knockout mutants (Δ*yejF*, Δ*yjjB*, Δ*yohN* and Δ*yrdB*) obtained from the Keio collection [34] confirmed *yohN* confers sensitivity to Co^{2+} and possibly Ni^{2+} (Fig. S7 in File S1). The growth of these strains was also tested in the presence of AMPs (Fig. 4B). While the MIC for pleurocidin was not affected by the presence of any of the four deletions, a modest but significant ($p<0.05$) increase in sensitivity was observed for all four deletion strains when challenged by magainin 2. When the experiment was repeated with LL-37, an AMP of human origin, three of the deletions rendered the bacteria more sensitive while deletion of *yrdB* had no effect.

The ontological profile related to each challenge offers another view of how closely related the response to each AMP is to each other. Here, instead of comparing individual genes on the basis of their identity, the comparison is based on the cellular component, biological process or molecular function and is less affected by redundancy or more subtle changes in response and consequently better reflects the fundamentals of the bacterial response. Ontological analysis, which employed a Benjamini-Hochberg method to control false discovery rate (FDR) and displays statistically overrepresented, differentially expressed genes in a graphical format according to their relationships in a hierarchical tree, was carried out on gene lists comprising the 200–250 most differentially expressed genes for each of the individual AMP treatments (Fig. S8–S13 in File S1) and for comparisons of up to three AMP treatments (Fig. 5; Fig. S8/S9 in File S1). The three AMPs derived from natural sources are suspected of acting on different cellular components. Indeed, comparing gene ontology (GO) term enrichment for cellular components (Fig. 5) showed a very different profile for each of magainin 2, buforin II and pleurocidin. Magainin 2 appears confined to affecting membrane components (Fig. 5; Fig. S10 in File S1) and had little effect on molecular functions or biological processes. Buforin II, in contrast, did not impact on any membrane components, instead focussing on components in the "cell" or "cell part" (Fig. 5; Fig. S11A in File S1) where 41% of the differentially expressed genes related to binding are found in the analysis of molecular function (Fig. S11B in File S1). Pleurocidin elicited responses both in membrane components and in the cell itself (Fig. 5; Fig. S12 in File S1) with biological processes, in particular polysaccharide and macromolecule metabolism and transport, impacted. This was reinforced by the finding that some 35 genes related to transporter activity were differentially expressed (Fig. S13 in File S1). These observations reinforce the view that AMPs impact on bacterial cells in distinct and AMP-specific ways. When the top 250 genes differentially expressed in response to challenge with D-LAK120-AP13 were analyzed, very few enriched pathways were found when biological processes were considered, with no enriched cellular components or molecular function identified. This indicates a non-specific response for this designed peptide notwithstanding its shared responses with buforin II observed above.

Discussion

The value of a combined approach

When taken together, the metabolomic, electron microscopy and transcript profiling analyses reveal a combination of generic and specific responses to challenge with AMPs that share many physicochemical features but that differ in their modes of action. All four peptides used were cationic, of similar lengths, and will

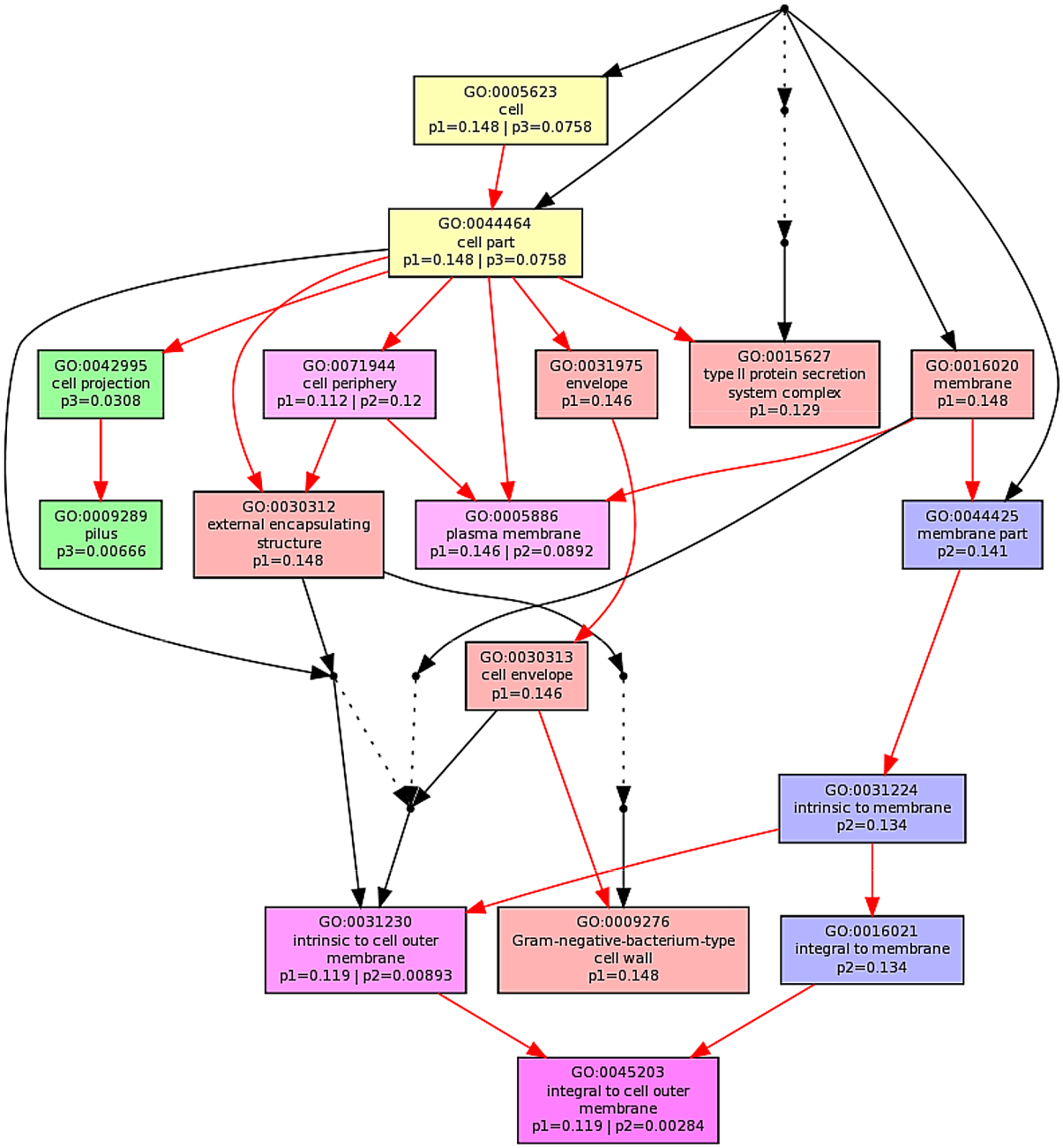


Figure 5. Multi GOEAST comparison of gene ontology (GO) terms relating to cellular component for differential gene responses in stationary phase *E. coli* NCTC 9001. Challenge was induced with sub-inhibitory concentrations of pleurocidin (red: p1), magainin 2 (blue: p2) and buforin II (green: p3). Red arrows represent relationships between two enriched GO terms, black arrows between enriched and un-enriched terms and black dashed arrows represent relationships between two un-enriched GO terms. Raw *p* values for GO terms have been adjusted using the Benjamini-Hochberg method allowing FDR<15%.
doi:10.1371/journal.ppat.1004104.g005

adopt conformations with secondary amphipathicity in the supposed target of the *E. coli* inner membrane. For the analytical techniques used, some strengths and weaknesses were identified, so underscoring the value of a combined approach. The electron micrographs provided compelling evidence of AMPs induction of manifestly different responses in *E. coli* challenged at both inhibitory and sub-inhibitory concentrations. The images however provide only circumstantial evidence as to the mechanism of action of each peptide. Instead, quantitative information or details of the molecular mechanisms involved are needed to pinpoint how each peptide operates. Transcript profiling provides a rich vein of information on the bacterial response. The individual gene products implicated have suggested a wide range of experiments that will illuminate further how bacteria attempt to fight off challenges posed by AMPs. Transcript profiling may also be more sensitive than the other approaches used since it alone was able to identify a significant response to buforin II which, even when administered at 250 μg/ml did not cause any perceived effect on either the internal or external cellular morphology or register a response as detected by ^{1}H HR-MAS NMR. The transcript

profiling method remains expensive however and the consumable costs per sample make its use in a high throughput manner unattractive. The NMR metabolomic technique has the advantage of having low per sample consumable costs which enables a much greater range of test conditions to be assessed. NMR metabolomics is also highly reproducible and provides quantitative information on this greater number of test conditions. It would therefore be attractive to consider whether it could be used as a standalone method for interrogating bacterial responses to challenge. In the present study however, while both generic and specific changes in metabolites were identified in response to AMP challenge, generic changes may appear overestimated when compared with the information provided by transcript profiling or micrographs. This may be due to common metabolic pathways underpinning a series of distinct bacterial responses and a much larger scale investigation, with a larger panel of both distinct and more closely related AMPs is now warranted. This would allow greater weight to be afforded to certain key metabolites, known to be altered in response to a given class of AMP with known influence on bacterial stress responses.

Life and death at the membrane?

This study investigates whether studying bacterial responses, when challenged with carefully defined sub-lethal concentrations of antibiotic, provides a detailed systems wide view of the mechanism of action. The mechanism of action of cationic amphipathic helix forming antimicrobial peptides has received considerable attention in the past two decades with much work focussed on the pore forming activity of magainin 2 and related peptides [35]. Considered an archetypal pore forming peptide, there is nevertheless evidence that for at least one microbial target, *Saccharomyces cerevisiae*, magainin 2 can enter the cell and interfere with DNA integrity [36] while pore forming activity that causes graded dye release is linked to a mechanism that involves translocation of the peptide across the membrane [37]. Finally, MD simulations have shown that magainin-H2, when forming a disordered toroidal pore does indeed translocate to the internal leaflet of the membrane [38]. Set against these studies are a range of data on the structurally and physico-chemically related, but considerably more potent, pleurocidin which is known to have pore forming activity [21] but is also capable of entering bacteria to interfere with the synthesis of macromolecules [22]. We have recently solved the high resolution structures of both magainin 2 and pleurocidin in the anionic detergent SDS (PDB entries 2LSA and 2LS9 respectively) and found similar regions of flexibility around the glycine residues in the middle section of the sequence (Gly 13/18 – magainin 2; Gly 13/17 – pleurocidin). Only in the membranes that most closely mimic the inner membrane of Gram-negative bacteria are any differences between the two peptides observed; here pleurocidin adopts a notably more disordered conformation under these conditions [18]. The more disordered conformation of pleurocidin in the *E. coli* target membrane may be related to possible pore formation [39] or the proposed intracellular targeting strategy [22] which, in both cases, would serve to boost its potency.

Previous 'omics based studies comparing AMPs action in Gram-positive bacterial species found that there was very little overlap in response between *Streptococcus pneumoniae* that had been challenged with each of three rather different antimicrobial peptides [11], while two earlier studies [9,10], which focussed on peptides with the plasma membrane as a presumed common target, found rather more overlap. We therefore decided to test whether a more holistic approach would succeed in discriminating between the different modes of actions of magainin 2 and pleurocidin and place their differing membrane activities in a wider context, enabling a more sophisticated understanding of their respective mechanisms of action while explaining the greater potency of pleurocidin. In the present study, the combined approach was readily capable of distinguishing pleurocidin and magainin 2 on the basis of the bacterial responses observed in their metabolomic and transcript profiles with electron micrographs bringing these differences into sharp relief. Despite the shared physicochemical properties and conformational propensities of the two peptides and presumed initial target of the bacterial inner membrane, transcript profiling identified only 19 genes whose differential expression was common to both AMP challenges, with differential expression of some 399 genes being a specific response to either pleurocidin or magainin 2. The *E. coli* response to AMP challenge is therefore highly adaptable and is most sensitive to the differing bactericidal strategies of each peptide. Large scale changes in the internal morphology of *E. coli*, following challenge with sub-inhibitory concentrations of each AMP, provides circumstantial evidence that both magainin 2 and pleurocidin can enter Gram-negative bacteria, with the more profound effects of pleurocidin suggesting a greater proficiency. Improvements in imaging technologies and labelling techniques may open the way, in future, for the more precise localisation of both peptides but it is apparent that a simple description of AMP bactericidal mechanisms that rests solely on studying the membrane interaction in isolation is inadequate. This is particularly relevant for the goal of increasing potency.

We have also studied the structural properties of buforin II which is considered to operate via an intracellular targeting strategy [18]. Buforin II has a greater affinity for nucleic acids, has a greater nominal charge at +7 and is less hydrophobic. The proline kink in buforin II is known to be crucial for enabling translocation into the *E. coli* cytosol [24]. Notably, in all membranes that we have studied, the peptide adopts an extended helical conformation, rather than one rich in α-helix, and has only barely detectable antibacterial activity against planktonic *E. coli* cultures [18]. We therefore included buforin II in the present study since we hypothesised that the bacterial response to this peptide would highlight responses to pleurocidin that are related to an intracellular targeting strategy. Neither the NMR metabolomic nor electron micrograph studies though identified a strong response to even very elevated concentrations of this peptide; consistent with our previous work which identified only a very weak effect against planktonic cultures of either *E. coli* or *P. aeruginosa* [18]. Nevertheless, a large number of significantly differentially expressed genes in response to buforin II challenge were detected by transcript profiling. While around 64 differentially expressed genes were detected in common to challenge with buforin II and pleurocidin, 33 differentially expressed genes were common to buforin II and magainin 2 with a further 534 differentially expressed genes identified that were not affected by either magainin 2 or pleurocidin. Only six differentially expressed genes were identified as a common response to these three AMPs. This further emphasises the plasticity of the *E. coli* response and indicates that bacteria have a large repertoire of responses to challenges.

Considering the ontology of the differentially expressed genes can suggest how each individual AMP operates but, when used in comparison, as here, the relative importance of the properties of each AMP is revealed and supported the view that these three peptides adopt distinct bactericidal strategies. The ontological profiles reveal near orthogonal changes in transcript profiles following sub-lethal challenge with the three different AMPs of natural origin. Comparison of GO terms with existing paradigms for the mode of action of each AMP supports the view that the

present, combined approach faithfully reveals the mechanism of action, notwithstanding the extra detail that identifies a range of effects that may contribute to bacterial cell death. In particular, the identification of eight GO terms linked to membranes supports the established view that magainin 2 largely acts on the plasma membrane of Gram-negative bacteria. In contrast, within the top 200 differentially expressed genes, no membrane GO terms were linked to the action of buforin II which is considered to seek intracellular targets. This is further supported by the distribution of GO terms since the effect on binding and a host of biosynthetic pathways is acute. For pleurocidin, where multiple bactericidal mechanisms have been proposed, there is substantial overlap between the cellular component GO terms with those affected by magainin 2. This indicates that the bacterial membrane is indeed a common target. However, in contrast with magainin 2, pleurocidin impacts on a large number of intracellular biological process, in particular macromolecule metabolic and transport processes. This strongly indicates a multifaceted antibacterial strategy underpins the high antibacterial potency of this AMP.

Can understanding the bacterial response be exploited to improve AMP potency?

The high plasticity of the bacterial response to AMP challenge suggests that deletion of one gene is unlikely to have a great impact on sensitivity. This view is supported by the study of mutants identified by mapping metabolite changes with the greatest pathway impact onto their respective pathways and further work will be required to more effectively disrupt such pathways in order to identify any relationship with sensitivity to AMPs.

Nevertheless, six gene products were identified that were significantly and uniformly affected by the three AMPs derived from natural sources. Of these six genes, two were down-regulated; *yrdB* an anionic 85 amino acid hypothetical protein and *yejF*. The *yejF* gene codes for the ATPase in the ABC transporter YejABEF which, when mutated, confers resistance to microcin C [33]. The speculated role of YejABEF as a nickel transporter has been questioned as it is phylogenetically distant from other oligopeptide transporters [33]. However, since *yejF* is down-regulated in the present study in response to all three peptides obtained from natural sources and its deletion renders *E. coli* more sensitive to both magainin 2 and LL-37, this behaviour does support the earlier finding that the activity of this protein can have a considerable effect on peptide antibiotic potency. Indeed, while mutations in *yejABEF* confer resistance to microcin C in *E. coli*, deletion of *yejF* in *Salmonella enterica* increased sensitivity to AMPs, including both human beta defensins 1 and 2 (hBD-1 and hBD-2) [40].

Of the four genes that are up-regulated, *cysE* and *manA* are widely distributed amongst taxa, including animals, making them less attractive as an antibiotic target. In contrast, with a distribution that is concentrated in *Enterobacteriaceae* and with yet to be tested functions, *yohN and yjjB* might be more attractive targets for further investigation and possible targets for adjuvants that could boost the potency of the host innate immune response. Deletion of these genes caused a significant but only modest increase in sensitivity to magainin 2 and LL-37 while the potency of pleurocidin was unaffected. These results show that the combined systems approach is indeed capable of identifying genes that regulate resistance/sensitivity in *E. coli* but that the large number of potentially differentially expressed genes at the disposal of such bacteria will mitigate the effect that silencing one gene product may have.

Finally, we were interested to contrast the expected results for the three peptides representing naturally occurring AMPs with the bacterial response to a peptide, D-LAK120-AP13, which was composed of D-amino acids only. D-LAK120-AP13 was designed in an attempt to circumvent the effect of proteases secreted by target pathogens, and incorporate structural features, including high cationicity and propensity for adopting α-helix rich conformation [41] - and hence inserting into and disordering the *E. coli* inner membrane - and a proline kink, affording conformational flexibility [20] that facilitates penetration into bacteria [23,24]. The robust and potent effect of this peptide against *E. coli* was evident with a significant metabolomic response even at very low peptide concentrations. Circumstantial evidence for the ability to penetrate within bacterial cells was shown by transmission electron microscopy, with the most profound changes due to challenge with any of the four AMPs observed, and transcript profiling. Again underlining the plasticity of the *E. coli* response, transcript profiling identifies a further 390 differentially expressed genes that were uniquely affected by D-LAK120-AP13 although, interestingly, there is considerable degree of overlap with the response to buforin II with 192 differentially expressed genes in common. These two peptides have a greater nominal cationic charge in solution at neutral pH than either pleurocidin or magainin 2 and both incorporate a proline induced kink in the secondary amphipathic conformation. Taken together, the data support highly effective entry of D-LAK120-AP13 into Gram-negative bacterial cells and it is this that may underpin its high antibacterial potency.

With four distinct but physicochemically related AMPs now tested by an integrated systems biology approach, a total of at least 1342 differentially expressed genes ($p \leq 0.05$) have been identified as being potential tools that can be manipulated by the bacteria to overcome AMP challenge. This is equivalent to between 24 and 32% of the total *E. coli* genome and suggests, with more structurally diverse AMPs yet to be tested, that bacteria have a wide variety of means of overcoming AMP challenges. Understanding these responses enables both the mode of action of AMPs to be elucidated as well as suggesting strategies to overcome these defences. The approach may find generic applicability in the study of antibiotic-bacteria arms races.

Materials and Methods

Materials

The peptides (Table 1) were all amidated at the C-terminus and were purchased from Pepceuticals Ltd (Nottingham, UK) as desalted grade or synthesised in house (D-LAK120-AP13) and were further purified using water/acetonitrile gradients using a Waters SymmetryPrep C8, 7 μm, 19×300 mm column.

Bacterial culture and challenge

Cultures of *Escherichia coli* NCTC 9001, a strain isolated from a patient with cystitis, were grown overnight in Mueller-Hinton broth (MH) at 37°C. Once the OD_{620} reached ≈1.0, 1 ml aliquots of bacterial suspension were transferred into 1.5 ml microcentrifuge tubes and aqueous solutions of peptides - magainin 2, buforin II, pleurocidin and D-LAK120-AP13 were added at the following concentrations: 250 μg/ml, 125 μg/ml, 62.5 μg/ml, 15.6 μg/ml, 3.9 μg/ml and incubated for 30 min at 37°C. In order to be able to monitor the microbial recovery and growth, 10 μl of each suspension was sampled in 190 μl fresh medium onto a 96-well microplate. The OD_{620} was measured at time 0 and after 4 h of incubation at 37°C. The microcentrifuge tubes were centrifuged at 5000× *g* for 5 min and the bacterial pellets were snap frozen in liquid nitrogen, lyophilised and kept at −20°C until further use. Pellets from triplicate tubes were combined for

subsequent HR-MAS analysis. Each challenge was independently repeated nine times.

HR-MAS NMR

High-resolution magic angle spinning (HR-MAS) experiments were performed on a Bruker Avance 400 MHz spectrometer equipped with a 4 mm $^{1}H/^{13}C$ HR-MAS probe. The lyophilised cell pellets were thawed at room temperature, transferred to an NMR rotor inserts and rehydrated with 30 µl of D_2O 2 hours before the acquisition. 1D spectra were recorded at a constant temperature of 310 K with magic angle spinning applied at 5 kHz. 1D ^{1}H spectra were recorded using a standard cpmgpr1d spin echo pulse (cpmgpr; Bruker) with water presaturation during recycle delay of 1 second and a total of 128 scans were acquired. The spectral width was 16.02 ppm and ^{1}H 90 pulse length was 7.81 µsec. The free induction decay was multiplied with an exponential function corresponding to a line broadening of 0.3 Hz. Phase correction was performed manually and automatic baseline correction was applied. A total of 120 samples were analysed with between 6 and 13 samples per treated condition and 17 control samples (no AMP treatment). A number of 2D experiments were run to facilitate identification of the compounds: homonuclear J-resolved 2D correlation with presaturation during relaxation delay using gradients (J-Res; jresgpprqf), $^{1}H/^{13}C$ correlation via direct inept transfer, phase sensitive using states, with decoupling during acquisition (HSQC 13C; AA-hsqcwg-13C), 2D homonuclear shift correlation with presaturation during relaxation delay (COSY; cosyprqf) all acquired using standard Bruker pulse sequences. Spectra were Fourier transformed, manually phase and automatically baseline corrected and calibrated with 2,2,3,3-D4-3-(Trimethylsilyl) propionic acid sodium salt (TMSP-2,2,3,3-D4) with reference signal at 0 ppm.

Assignment

Resonances were assigned based on J-couplings partners revealed by COSY, multiplicities derived from J-Res, statistical correlation spectroscopy (STOCSY) [42] and both ^{1}H and ^{13}C chemical shifts with reference to the *E. coli* metabolome database [43].

Multivariate data analysis

Spectra were analysed by principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) using software developed in our laboratory for a previous study [44] incorporating the nonlinear iterative partial least squares (NIPALS) algorithm [45]. First, the spectra were aligned to the reference peak and spectral regions such as water and reference peak (4.8 ppm and 0 ppm, respectively) and regions of no interest and/or no spectral information were removed. Spectra were then normalised using probabilistic quotient normalization (PQN) [46] and autoscaled but not bucketed. Cross-validation was performed where 66% of the samples were used as a training set and the remaining 33% as a test set, ensuring that the number of samples in the test set was proportional to the total number of samples from each class, and that at least one sample from each class was present in the test set. To choose the number of components for the model, a leave-one-out cross-validation was carried out on the samples in the training set, and the F1-score used to choose the number of components, with the additional constraint to use a maximum of 10 components. This double cross-validation was repeated 2000 times with randomly chosen samples in the training and test set to prevent bias due to the choice of training or test set. This leads to 3×2000 models (in the supplementary information, each of these models leads to a point on the scores plot, but loadings and weights are presented as averages over all these models). Finally, this procedure was repeated with randomly generated class assignments to provide a reference value for Q^2. The chosen number of components minus one was then used as an OPLS filter and a PLS-DA analysis with two components was carried out on the filtered data to yield one predictive and one orthogonal component. The Q^2 value was calculated as $Q^2 = 1-(\mathrm{PRESS}/\mathrm{TSS})$ where PRESS is the sum of squared differences between the known and predicted classes, and TSS is the sum of squared differences between the known classes and their average (= the total variance). Q^2 thus gives a measure of the goodness of fit after cross validation, and although it is generally considered to be "good" when its value is higher than 0.5 [47,48] we have compared it to a reference value by computing Q^2 for models where the classes were assigned randomly [47,48]. In each case, genuine or permutated class assignments, the Q^2 value quoted is the mean of all models. Back-scaled loadings plots [49] were used to identify resonances with high variance and high weight, therefore the discriminating resonances, and verified against the peak intensity of the original spectra after PQN normalisation. Freely available MultiExperiment Viewer (MeV) which is a part of the TM4 Microarray Software Suite [50] was used for hierarchical cluster (HCL) analysis and generation of heatmaps. Euclidian distance algorithm was used to compute the differences between two gene expression levels (metabolite level changes) and the average linkage method was used to define the distances.

Scanning and transmission electron microscopy

Both SEM and TEM were used to examine the structural changes in bacteria induced by AMPs. Samples for the imaging were prepared in parallel with the samples used for HR-MAS NMR and hence represent bacteria in stationary phase. For SEM, the pellet obtained after centrifugation was fixed in 25 µl of 2.5% (v/v) glutaraldehyde in 0.2 M sodium cacodylate buffer and kept at 4°C until further use. In 24-well tissue culture plates 20 µl aliquots of vortexed bacterial pellet was smeared on 12 mm round poly-L-lysine (BD Biosciences, Bedford) cover slips with adjacent chambers filled with sufficient amount of 0.2 M sodium cacodylate to prevent drying of the slides and kept in a hydration chamber for 2 h. Cover slips were then washed with 0.2 M sodium cacodylate buffer followed by rinsing with 30%, 70%, 100%, 100%, and 100% ethanol and incubating for 10 min between each wash. Hexamethyldisilazane (HMDS) was used for drying of the specimen by washing cover slips in 50/50 100% ethanol/HMDS for 10 min followed by the final wash in HMDS for 10 min. The coverslips with dehydrated cells were mounted on the specimen stubs and sputter coated with gold. Micrographs were acquired with FEI Quanta 200F FEG scanning electron microscope. Bacterial pellets for TEM processing were prepared as described above. Cells were pelleted by centrifugation and the pellet was post fixed in 1% osmium tetroxide in 0.1 M phosphate buffer for 60 min an RT. The pellet was dehydrated by exposure to a graded series of ethanol (10%, 70% for 10 min each) followed by four washes in 100% ethanol for 15 min each. Next, the pellet was subjected to two washes in propylene oxide, 10 min each. Tubes containing pellets were constantly rotated during the washes and the following procedures and the washes were performed in the fume hood. The supernatant was removed and the pellet placed into a mixture of 50% resin and propylene oxide for 90 min and transferred to 100% resin overnight before polymerisation at 60°C for 24 hours. The resin blocks were sectioned with Leica Ultra-cut ultramicrotome to semi-thick sections (0.75 µm–2 µm) and stained with toluidine blue and used to determine the areas for thin

sectioning (90 nm). The sections were then placed onto 150 mesh copper grids coated with pioloform support film. Grids were then stained with uranyl acetate and lead citrate before viewing on Hitachi H7600 transmission electron microscope. For both techniques, around 15 images were taken for each treatment. The following magnifications were used and images were selected that are representative of the effect observed: 700×, 5000×, 12000×, 25000×, 70000×.

GeneChips

GeneChip experiments were performed using the Affymetrix (Santa Clara, CA) *E. coli* Genome 2.0 Array with effective, response inducing, sub-MIC AMP concentrations determined from the HR-MAS metabolomic study; pleurocidin 62.5 μg/ml, buforin II 250 μg/ml, magainin 2 125 μg/ml and D-LAK120-AP13 15.6 μg/ml. Each array includes approximately 10,000 probe sets for all 20,366 genes present in four strains of *E. coli* over the entire open reading frame (ORF); K12 (MG1655 laboratory strain), CFT073 (uropathogenic), 0157:H7-EDL953 (enteropathogenic) and O157:H7-Sakai (enteropathogenic). RNA was extracted using RiboPure and enriched using MICROB*Express* Bacterial mRNA Enrichment Kit after the DNA digestion step (Life Technologies, Paisley, UK) At each step the quality of RNA was assessed using Pico100 (Picodrop Ltd, Hinxton, UK). cDNA was synthesized from mRNA and purified using Qiagen MinElute PCR (Qiagen, Manchester, UK). cDNA was then fragmented and labeled using terminal transferase and biotinylated Affymerix GeneChip labelling reagent according to the manufacturer's instructions. Fragmentation and labeling were assessed with the 2100 Bioanalyzer (Agilent Technologies, Wokingham, UK) to obtain the size distribution and yield. cDNA was kept at −80°C until microarray hybridization. Hybridization of the target to the GeneChip was prepared according the standard Prokaryotic Target Hybridisation protocol according to the manufacturer's instructions. The efficiency of the hybridization step was assessed by examining hybridization of Poly-A controls provided for the Affymetrix GeneChip. Arrays were scanned on an Affymetrix GCS3000 microarray system and image acquisition, quantification and data analysis were performed using Affymetrix Command and Expression Console Software. Data were normalized using the Robust Multi-array Average (RMA) algorithm built into Expression Console. Pre-selection of gene lists for each treatment was performed using Qlucore Omics Explorer (Qlucore AB, Lund, Sweden). First, ANOVA across all samples identified the twenty most differentially expressed genes according to each replicated treatment. These were then assessed by principal component analysis (Fig. S6 in File S1) to confirm that independently replicated experiments produced consistent results. Signal intensities for gene expression were then averaged across technical duplicates/triplicates and log transformed. For the gene annotation enrichment analysis, differentially expressed genes in treatment versus control samples were selected by a paired, homoscedastic t-test with a significance cutoff of $p<0.05$ and lists for the four AMP treatments were then compared using Venny [51]. Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-1703. To better understand the differences between the effects of the four treatments, significance thresholds that identified the approximate top 200–250 differentially expressed genes were selected; $p\leq0.0184$ for buforin II and D-LAK120-AP13, $p\leq0.0425$ for pleurocidin and $p\leq0.078$ for magainin 2. These lists were analyzed using the GOEAST Gene Ontology Enrichment Analysis Software Toolkit where the Benjamini-Hochberg option was selected allowing an FDR up to 15% [52]. Discriminating metabolite changes, identified from HR-MAS NMR, were then mapped onto the KEGG pathway using BioCyc Omics Data Analysis [53] and genes related to given metabolic pathway checked against consistently differentially expressed genes, whether or not they had passed the significance test described above.

Multiparameter viability assays

In order to assess the functionality and cellular integrity of bacteria we used the following viability assays: membrane potential assay, esterase activity assay and BacLight Live-Dead stain for microscopy [26].

As previously, *E. coli* NCTC 9001 were grown from glycerol stocks in Muller-Hinton broth overnight at 37°C without shaking until an OD_{620} of 1.0 was reached. 1 ml aliquots of culture were challenged for 30 min with four peptides at and below the threshold concentrations established with NMR. Cells were then harvested by centrifugation at 5,000× g for 5 min and washed in 50 mM phosphate buffer (pH 7.0). For BacLight Live/Dead stain cells were diluted to 4×10^8 CFU/ml, whereas for the remaining assays cells were diluted to 2×10^8 CFU/ml. All experiments were performed at room temperature. Negative controls were obtained either by treatment with 70% isopropanol for 10 min and removed by centrifugation at 5,000× g for 5 min and re-suspension in PBS, or by heat killing at 85°C for 10 min on a heat block. Assays were performed in black, flat bottom, 96-well plates and read on a Synergy HT multi-mode microplate reader (BioTek, Winooski, VT)

Membrane potential

25 mg of dye $DiBAC_4$ (Anaspec, Fremont, CA) was reconstituted in 2.42 ml ethanol to obtain a 20 mM stock solution which was stored at −20°C. The stock was diluted further with water to working concentration of 12.5 μM immediately before use. 20 μl of 12.5 μM dye was added to a 96-well plate, covered by 180 μl bacterial suspension in PBS and mixed. The plate was incubated in the dark for 5 minutes and fluorescence emission was measured (excitation 485 nm, emission 535 nm). Since membrane damage leads to higher fluorescence intensity, values were background corrected and expressed as a reciprocal before being normalised with untreated cells defined as being 100% and isopropanol treated cells defined as 0%.

Esterase activity

5 mg of esterase substrate 5,6-carboxyfluorescein diacetate (CFDA) was dissolved in 1.086 ml dimethyl sulfoxide (DMSO) to obtain 10 mM stock kept at −20°C. Stock was diluted 40× in water immediately before use to obtain working concentration of 250 μM, which was pre-aliquoted to a 96-well plate. 180 μl of bacterial suspension in PBS was added to the plate and mixed with the detection solution. The plate was incubated in dark for 30 minutes with occasional shaking and fluorescence emission measured (excitation 485 nm, emission 535 nm).

LIVE/DEAD BacLight

LIVE/DEAD BacLight kit (Life Technologies, Paisley, UK) was used to measure membrane integrity. Harvested cells were reconstituted with saline and 3 μl of the dye mixture (1.5 μl of SYTO9 (3.34 mM) and 1.5 μl of propidium iodine (20 mM)) was added to each 1 ml of bacterial suspension and mixed. Tubes were incubated for 15 minutes in the dark with occasional shaking and fixed with 20% paraformaldehyde (PFA) and kept at 4°C. Specimens were viewed on an Olympus BX60 microscope fitted

with an Andor Ultrahigh-resolution CCD setup. A ×20 oil immersion lens was used to obtain a 200 µm field width. Excitation and emission filters were 480/520 nm and 515/560 nm respectively.

MIC testing

Parent strain BW25113 and Keio knockout strains [34] for Δ*yejF*, Δ*yjjB*, Δ*yohN*, and Δ*yrdB* were obtained from the Coli Genetic Stock Center (Yale University, New Haven, CT). The activities of the peptides were assessed in planktonic suspension in polypropylene 96 well plates (Greiner Bio-one, Frickhausen, Germany) according to a modified broth dilution assay (54). Bacteria were grown without shaking in 50 ml Mueller-Hinton (MH) broth at 37°C. Peptides (pleurocidin, magainin 2 and LL-37) were tested in duplicates with two rows allocated for each peptide. In each of columns 2–11, 50 µl of MH broth was added under sterile conditions. In the first column of each row, 50 µl of 256 µg/ml stock peptide solutions, prepared in distilled water, were added and then the broth from the second column was pipetted into the first column and thoroughly mixed before being deposited again in the second column. This process was repeated throughout the tray providing a twofold dilution of peptide with each row. Bacteria with an A_{620} of 0.001 were then added to each well in volumes of 50 µl giving a further twofold dilution and a final volume of 100 µl per well. The final column was used either as sterility control (100 µl broth) or negative control (no peptide). Plates were incubated overnight at 37°C and the A_{620} read. Growth curves prepared from duplicates were fitted to determine the peptide concentration required to inhibit growth by 50% (MIC_{50}). The MIC_{50} quoted for each peptide (Fig. 4) is an average value from at least two independent repeats.

Supporting Information

File S1 Supplementary figures and table.
(PDF)

Acknowledgments

Dr. Matt Arno & Dr. Estibaliz Aldecoa-Otalora Astarloa provided assistance and access to the KCL Genomics Centre while the KCL Centre for Ultrastructural Imaging supported electron microscopy. Dr. Andreas Nocker offered valuable advice. LL-37 peptide was a kind gift from Giulia Armato and Dr. Sukhi Bansal.

Author Contributions

Conceived and designed the experiments: JK MM KDB AJM. Performed the experiments: JK LSV GBR NR SBTAA GK. Analyzed the data: JK LSV AJM. Contributed reagents/materials/analysis tools: LSV MM KDB AJM. Wrote the paper: JK LSV GBR KDB AJM.

References

1. Steiner H, Hultmark D, Engstrom A, Bennich H, Boman HG (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. Nature 292: 246–248.
2. Zasloff M (1987) Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. Proc Natl Acad Sci USA 84: 5449–5453.
3. Ganz T, Selsted ME, Lehrer RI (1990) Defensins. Eur J Haematol 44: 1–8.
4. Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nature Rev Microbiol 3: 238–250.
5. Hancock REW, Sahl H-G (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nature Biotech 24: 1551–1557.
6. Peschel A, Sahl H-G (2006) The co-evolution of host cationic antimicrobial peptides and microbial resistance. Nature Rev Microbiol 4: 529–536.
7. Hale JDF, Hancock REW (2007) Alternative mechanisms of action of cationic antimicrobial peptides on bacteria. Expert Rev Anti Infect Ther 5(6): 951–959.
8. Wecke T, Mascher T, (2011) Antibiotic research in the age of omics: from expression profiles to interspecies communication. J Antimicrob Chemother 66(12): 2689–2704.
9. Pietiäinen M, Gardemeister M, Mecklin M, Leskelä S, Sarvas M, et al. (2005) Cationic antimicrobial peptides elicit a complex stress response in *Bacillus subtilis* that involves ECF-type sigma factors and two-component signal transduction systems. Microbiology 151: 1577–1592.
10. Pietiäinen M, François P, Hyyryläinen H-L, Tangomo M, Sass M, et al (2009) Transcriptome analysis of the responses of *Staphylococcus aureus* to antimicrobial peptides and characterization of the roles of vraDE and vraSR in antimicrobial resistance. BMC Genomics 10: 429
11. Majchrzykiewicz JA, Kuipers OP, Bijlsma JJE (2010) Generic and specific adaptive responses of *Streptococcus pneumonia* to challenge with three distinct antimicrobial peptides, bacitracin, LL-37, and nisin. Antimicrob. Agents Chemother 54(1): 440–451
12. Hong RW, Shchepetov M, Weiser JN, Axelsen PH (2003) Transcriptional profile of the *Escherichia coli* response to the antimicrobial insect peptide cecropin A. Antimicrob Agents Chemother 47(1): 1–6.
13. Tomasinsig L, Scocchi M, Mettulio R, Zanetti M (2004) Genome wide transcriptional profiling of the *Escherichia coli* response to a proline-rich antimicrobial peptide. Antimicrob Agents Chemother 48(9): 3260–3267.
14. Kruse T, Christensen B, Raventós D, Nielsen AK, Nielsen JD, et al. (2009) Transcriptional profile of *Escherichia coli* in response to novispirin G10. Int J Pept Res Ther 15: 17–24.
15. Mason AJ, Chotimah INH, Bertani P, Bechinger B (2006) A spectroscopic study of the membrane interaction of the antimicrobial peptide Pleurocidin. Mol Membr Biol 23(2): 185–194.
16. Mason AJ, Marquette A, Bechinger B (2007) Zwitterionic phospholipids and sterols modulate antimicrobial peptide induced membrane destabilisation. Biophys J 93: 4289–4299
17. Mason AJ, Bertani P, Moulay G, Marquette A, Perrone B, et al (2007) The membrane interaction of chrysophsin1, a histidine rich antimicrobial peptide from red sea bream. Biochemistry 46(51): 15175–15187
18. Lan Y, Yan Y, Kozlowska J, Lam JKW, Drake AF, et al. (2010) Structural contributions to the intracellular targeting strategy of antimicrobial peptides. Biochim Biophys Acta 1798: 1934–1943.
19. Mason AJ, Moussaoui W, Abdelrahman T, Boukhari A, Bertani P, et al (2009) Structural determinants of antimicrobial and antiplasmodial activity and selectivity in histidine rich amphipathic cationic peptides. J Biol Chem 284: 119–133
20. Vermeer LS, Lan Y, Abbate V, Ruh E, Bui TT, et al (2012) Conformational flexibility determines selectivity and antibacterial, antiplasmodial and anticancer potency of cationic α-helical peptides. J Biol Chem 287: 34120–34133.
21. Saint N, Cadiou H, Bessin Y, Molle G (2002) Antibacterial peptide pleurocidin forms ion channels in planar lipid bilayers. Biochim Biophys Acta 1564: 359–364.
22. Patrzykat A, Friedrich CL, Zhang L, Mendoza V, Hancock REW (2002) Sublethal concentrations of pleurocidin-derived antimicrobial peptides inhibit macromolecular synthesis in *Escherichia coli*. Antimicrob Agents Chemother 46: 605–614.
23. Park CB, Kim HS, Kim SC (1998) Mechanism of action of the antimicrobial peptide buforin II: Buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. Biochem Biophys Res Comm 244: 253–257.
24. Park CB, Yi K-S, Matsuzaki K, Kim MS, Kim SC (2000) Structure-activity analysis of buforin II, a histone H2A-derived antimicrobial peptide: the proline hinge is responsible for the cell penetrating ability of buforin II. Proc Natl Acad Sci USA 97: 8245–8250.
25. Wiegand I, Hilpert K, Hancock RE (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature Protocols 3: 163–175.
26. Nocker A, Caspers M, Esveld-Amanatidou A, van der Vosen J, Schuren F, et al (2011) Multiparameter viability assay for stress profiling applied to the food pathogen *Listeria monocytogenes* F2365. Appl Environ Microbiol 77(18): 6433–6440.
27. Xia J, Mandal R, Sinelnikov I, Broadhurst D, Wishart DS (2012) MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. Nucl Acids Res 40: W127–W133
28. Xia J, Psychogios N, Young N, Wishart DS, (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucl Acids Res 37: W652–W660.
29. McBroom AJ, Kuehn MJ (2007) Release of outer membrane vesicle by Gram-negative bacteria is a novel envelope stress response. Mol Microbiol 63(2): 545–558.
30. Blattner FR, Plunkett G, Bloch CA, Perna NY, Burland V, et al (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453–1462.
31. Welch RA, Burland V, Plunkett G, Redford P, Roesch p, et al (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci USA 99(26): 17020–17024.

32. Blériot C, Effantin G, Lagarde F, Mandrand-Berthelot M-A, Rodrigue A (2011) RcnB is a periplasmic protein essential for maintaining intracellular Ni and Co concentrations in *Escherichia coli*. J Bacteriol 193: 3785–3793.
33. Novikova M, Metlitskaya A, Datsenko K, Kazakov T, Kazakov A, et al (2007) *The Escherichia coli* Yej transporter is required for the uptake of translation inhibitor microcin C. J Bacteriol 189(22): 8361–8365.
34. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2: 2006.0008
35. Ludtke SJ, He K, Heller WT, Harroun TA, Yang L, et al. (1996) Membrane pores induced by magainin. Biochemistry 35: 13723–13728.
36. Morton CO, Hayes A, Wilson M, Rash BM, Oliver SG, et al. (2007) Global phenotype screening and transcript analysis outlines the inhibitory mode(s) of action of two amphibian-derived, α-helical, cationic peptides on *Saccharomyces cerevisiae*. Antimicrob Agents Chemother 51: 3948–3959.
37. Gregory SM, Pokorny A, Almeida PFF (2009) Magainin 2 revisited: a test of the Quantitative Model for the all-or-none permeabilization of phospholipid vesicles. Biophys J 96: 116–131.
38. Leontiadou H, Mark AE, Marrink SJ (2006) Antimicrobial peptides in action. J Am Chem Soc 128: 12156–12161.
39. Rzepiela AJ, Sengupta D, Goga N, Marrink SJ (2010) Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions. Faraday Discuss 144: 431–443.
40. Eswarappa SM, Panguluri KK, Hensel M, Chakravortty D (2008) The yejABEF operon of *Salmonella* confers resistance to antimicrobial peptides and contributes to its virulence. Microbiology 154: 666–678.
41. Epand RF, Maloy WL, Ramamoorthy A, Epand RM (2010) Probing the "charge cluster mechanism" in amphipathic helical cationic antimicrobial peptides. Biochemistry 49: 4076–4084
42. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, et al (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. Anal Chem 77(5): 1282–1289.
43. Guo AC, Jewison T, Wilson M, Liu Y, Knox C, et al (2013) ECMDB: The *E. coli* metabolome database. Nucleic Acids Res 41(D1): D625–630.
44. Vermeer LS, Fruhwirth GO, Pandya P, Ng T, Mason AJ (2012) NMR metabolomics of MTLn3E breast cancer cells identifies a role for CXCR4 in lipid and choline regulation. J Proteome Res 11: 2996–3003
45. Andersson M (2009) A comparison of nine PLS1 algorithms. Journal of Chemometrics 23: 518–529.
46. Dieterle F, Ross A, Schlotterbeck G, Senn H (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. Anal Chem 78: 4281–4290.
47. Szymanska E, Saccenti E, Smilde AK, Westerhuis JA (2012) Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 8: S3–S16
48. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, et al (2008) Assessment of PLSDA cross validation. Metabolomics 4: 81–9
49. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, et al. (2005) Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in ^{1}H NMR spectroscopic metabonomic studies. Anal Chem 77: 517–26.
50. Saeed AI, Sharov V, White J, Li J, Liang W, et al (2003) Tm4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2): 374–378.
51. Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/ Accessed 24th March 2014
52. Zheng Q, Wang X-J (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res 36: W358–W363.
53. Available: http://ecocyc.org/PToolsWebsiteHowto.shtml#omicsDataAnalysis. Accessed 16 December 2013.